## GenCore version 5.1.6 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

January 7, 2004, 16:44:17; Search time 44 Seconds Run on:

(without alignments)

1215.701 Million cell updates/sec

US-10-088-872-2 Title:

1704

1 MKKMPLFSKSHKNPAEIVKI......FADEKNYLIKQIRDLKKTAP 337 Perfect score: Sequence:

BLOSUM62 Scoring table:

Gapop 10.0 , Gapext 0.5

1107863 seqs, 158726573 residues Searched:

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

A\_Geneseq\_19Jun03:\* Database :

1: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:\*

/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:\*

/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:\*

/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:\*

/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:\*

/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:\*

/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:\* 7:

/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:\*

9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:\*

10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:\* 11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:\*

/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:\*

12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:\*

13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:\*

/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:\*

15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:\*

16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:\* 17:

/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:\*

/SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:\* 19:

/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:\*

20:

/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:\* 21:

/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:\* 22:

/SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:\* 23:

/SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

## and is derived by analysis of the total score distribution.

### SUMMARIES

			ક				
Resi	.1+		Query				
	10.	Score		Length	DB	ID	Description
	·						
	1	1704	100.0	337	21	AAY94247	Human calcium bind
	2	1704	100.0	337	22	AAM39078	Human polypeptide
	3	1704	100.0	337	22	AAB82090	Human Acute Neuron
	4	1466	86.0	289	22	AAB94139	Human protein sequ
	5	1381	81.0	341	22	AAB48970	Human ANIC-BP (acu
	6	1381	81.0	496	22	AAE10858	Gal4-human ANIC-BP
	7	1381	81.0	552	22	AAE10859	LexA-human ANIC-BP
	8	1376	80.8	341	21	AAY94248	Mouse calcium bind
	9	1354	79.5	354	22	ABG23844	Novel human diagno
	10	1297.5	76.1	350	22	AAB20387	Human acute neuron
	11	1162	68.2	237	22	AAM40864	Human polypeptide
	12	1111	65.2	339	22	ABB60392	Drosophila melanog
	13	1109	65.1	339	21	AAY94249	Drosophila calcium
	14	1063.5	62.4	377	21	AAY94250	C. elegans yeast-l
	15	716.5	42.0	343	21	AAG45273	Arabidopsis thalia
	16	689.5	40.5	300	21	AAG23886	Arabidopsis thalia
	17	685.5	40.2	400	21	AAG51052	Arabidopsis thalia
	18	685.5	40.2	504	21	AAG51051	Arabidopsis thalia
	19	685	40.2	300	21	AAG30714	Arabidopsis thalia
	20	685	40.2	300	21	AAG45274	Arabidopsis thalia
	21	685	40.2	305	21	AAG30713	Arabidopsis thalia
	22	684.5	40.2	326	21	AAG51053	Arabidopsis thalia
	23	675.5	39.6	290	21	AAG23887	Arabidopsis thalia
• 5	24	671.5	39.4	345	21	AAG05089	Arabidopsis thalia
	25	638.5	37.5	320	21	AAG05090	Arabidopsis thalia
	26	539.5	31.7	213	21	AAG23888	Arabidopsis thalia
	27	533	31.3	213	21	AAG30715	Arabidopsis thalia Arabidopsis thalia
	28	533	31.3	213	21		Arabidopsis thalia
	29	478.5	28.1		21		Zea mays protein f
	30	467.5	27.4	154			Zea mays protein f
	31	453.5	26.6	148	21		Zea mays protein f
*	32	438.5	25.7			AAG41153 ABP02921	Human ORFX protein
	33	250.5	14.7				Novel human diagno
	34	241	14.1	639			Human ORF3054 prot
	35	227.5	13.4	135 383			Novel human diagno
	36	226.5	13.3 7.3				Peptide #3468 enco
	37	125	7.3				Human peptide enco
	38	125	6.9				Human novel polype
	39 40	117.5 114.5	6.7				Human ORFX ORF2117
		114.5	6.7				Human kinesin-rela
	41 42	113.5	6.7				Human kinesin moto
	42	113.5	6.7				Human partial kine
	44	113.3	6.6				P. falciparum telo
	45	111.5	6.5				Human RHAMM protei
	43	111.3	0.5	, 23			•

```
RESULT 1
     AAY94247 standard; protein; 337 AA.
AAY94247
ID
XX
     AAY94247;
AC
XX
                  (first entry)
     10-AUG-2000
DT
     Human calcium binding protein hCBP.
XX
DΕ
     Human; calcium binding protein; cancer; inflammation; CBP;
XX
     reproductive disorder; autoimmune disorder; developmental disorder;
ΚW
     seizure disorder; immune disorder; infection.
KW
 KW
 XX
      Homo sapiens.
 OS
 XX
      WO200029580-A1.
 PN
 XX
      25-MAY-2000.
 PD
 XX
                     99WO-US27027.
      12-NOV-1999;
 PF
 XX
                     98US-0190965.
      13-NOV-1998;
 PR
 XX
       (INCY-) INCYTE PHARM INC.
 PΑ
      Tang YT, Guegler KJ, Corley NC, Gorgone GA;
 XX
 PΙ
 XX
       WPI; 2000-387793/33.
  DR
       N-PSDB; AAA27332.
  DR
       Human hCBP protein, and the nucleic acid encoding it, useful for e.g.
  XX
       diagnosis, prevention and treatment of cancers, immune, developmental
  PT
  PT
       or reproductive disorders -
  PT
  XX
       Claim 1; Fig 1; 72pp; English.
  PS
       The present sequence is the human calcium binding protein hCBP. It
  XX
       was obtained by screening a coronary artery smooth muscle cDNA library,
  CC
       from which five overlapping nucleic acids were isolated, sequenced and
  CC
       expressed to give the protein. The protein and the gene encoding it are
  CC
       useful for the diagnosis and treatment of the following types of
  CC
       disorder: cancers (such as adenocarcinomas), reproductive disorders
  CC
        (such as infertility, ovulatory defects, endometriosis, disruptions of
  CC
        the oestrus and menstrual cycles, polycystic ovary syndrome and ovarian
  CC.
        hyperstimulation), autoimmune disorders (such as benign prostatic
   CC
        hyperplasia and prostatitis), developmental disorders (such as
   CC
        Cushing's syndrome, muscular dystrophy and gonadal dysgenesis),
   CC
        hereditary neuropathies, seizure disorders, immune disorders (such as
   CC
        AIDS, allergies, anaemia, asthma, atherosclerosis, cholecystitis, Crohn's
   CC
        disease, diabetes, Graves' disease, multiple sclerosis, psoriasis,
   CC
        rheumatoid arthritis, scleroderma, Sjogren's syndrome and ulcerative
   CC
        colitis), and viral, bacterial, fungal, parasitic, protozoal and
   CC
   CC
        helminthic infections.
   CC
   XX
```

337 AA;

Sequence

SQ

```
Length 337;
                      100.0%; Score 1704; DB 21;
                     100.0%; Pred. No. 1.3e-146;
 Query Match
 Best Local Similarity
                                                                 0:
                                                          Gaps
                                              Indels
                                                       0;
                           0; Mismatches
                                           0;
 Matches 337; Conservative
          1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEK\mathbf{E} 60
            QУ
          1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKB 60
DЪ
         61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
            Qy
         61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
Db
        121 AHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
            QУ
         121 AHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
Db
         181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
            Qу
         181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
Db
         241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPK 300
             Qу
         241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPK 300
 Db
         301 LIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
             Qy
         301 LIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
 Db
 RESULT 2
 AAM39078
     AAM39078 standard; Protein; 337 AA.
 TD
 XX
      AAM39078;
 AC
 XX
      22-OCT-2001 (first entry)
 DT
      Human polypeptide SEQ ID NO 2223.
 XX
  DE
      Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
  XX
      peripheral nervous system; neuropathy; central nervous system; CNS;
  ΚW
      Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
  KW
      amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
  KW
      chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
  KW
  ΚW
      leukaemia.
  KW
  XX
      Homo sapiens.
  OS
  XX
      WO200153312-A1.
  PN
  XX
       26-JUL-2001.
  PD
  XX
       26-DEC-2000; 2000WO-US34263.
  PF
  XX
       21-JAN-2000; 2000US-0488725.
  PR
       25-APR-2000; 2000US-0552317.
   PR
       09-JUL-2000; 2000US-0598042.
   PR
```

```
19-JUL-2000; 2000US-0620312.
PR
    03-AUG-2000; 2000US-0653450.
PR
    14-SEP-2000; 2000US-0662191.
PR
    19-OCT-2000; 2000US-0693036.
PR
    29-NOV-2000; 2000US-0727344.
PR
XX
     (HYSE-) HYSEQ INC.
PΑ
                                                       Ren F, Wang D;
                                              Qian XB,
                                      Ма Y,
XX
                     Asundi V, Chen R,
     Tang YT, Liu C,
                                                       Zhanq J;
                     Wehrman T, Xu C, Xue AJ,
                                               Yang Y,
PI
             Wang Z,
                      Goodrich R, Drmanac RT;
     Wang J,
PI
              Zhou P,
     Zhao QA,
PΙ
XX
     WPI; 2001-442253/47.
DR
     N-PSDB; AAI58234.
DR
     Novel nucleic acids and polypeptides, useful for treating disorders
XX
     such as central nervous system injuries -
PT
 PT
     Example 4; SEQ ID NO 2223; 10078pp; English.
 XX
 PS
     The invention relates to human nucleic acids (AAI57798-AAI61369) and
 XX
     the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC
     immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC
     in gene therapy. A composition containing a polypeptide or polynucleotide
 CC
     of the invention may be used to treat diseases of the peripheral nervous
 CC
      system, such as peripheral nervous injuries, peripheral neuropathy and
 CC
      localised neuropathies and central nervous system diseases, such as
 CC
      Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC
      lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC
      utilisation of the activities such as: Immune system suppression,
 CC
      Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC
      and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC
      assays for receptor activity, arthritis and inflammation, leukaemias and
 CC
  CC
      Note: The sequence data for this patent did not form part of the printed
  CC
  CC
      specification.
  CC
  XX
      Sequence
                 337 AA;
  SQ
                                                     Length 337;
                                  Score 1704; DB 22;
                          100.0%;
                                  Pred. No. 1.3e-146;
    Query Match
                          100.0%;
                                                                         0:
                                                                  Gaps
    Best Local Similarity
                                                     Indels
                                                 0;
                                0; Mismatches
    Matches 337; Conservative
             1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKE 60
               QУ
             1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKE 60
  Db
             61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
                QУ
             61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
   Db
            121 AHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
                Qу
            121 AHPHILFMLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
   Db
            181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
                QУ
```

```
181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
Db
         241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPK 300
             QУ
         241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPK 300
Db
         301 LIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
Qу
              301 LIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
Db
RESULT 3
AAB82090
     AAB82090 standard; Protein; 337 AA.
ID
XX
     AAB82090;
AC
XX
     26-JUN-2001 (first entry)
DТ
     Human Acute Neuronal Induced Calcium Binding Protein, ANIC-BP.
XX
DE
     Human; cerebroprotective; neuroprotective; vulnerary; vaccine;
XX
     gene therapy; Acute Neuronal Induced Calcium Binding Protein; ANIC-BP;
KW
     stroke; acute head trauma; multiple sclerosis; spinal cord injury.
KW
KW
 XX
    Homo sapiens.
 os
 XX
      WO200123552-A1.
 PN
 XX
      05-APR-2001.
 PD
 XX
      18-SEP-2000; 2000WO-EP09132.
 PF
 XX
                    99EP-0118848.
      24-SEP-1999;
 PR
 XX
      (MERE ) MERCK PATENT GMBH.
 PΑ
 XX
      Den Daas I, Duecker K;
 _{\mathrm{PI}}
 XX
      WPI; 2001-308142/32.
 DR
      N-PSDB; AAF86462.
 DR
      Novel human acute neuronal induced calcium binding polypeptide, and
 XX
      polynucleotides encoding them useful for diagnosing or treating stroke,
 PT
      acute head trauma, multiple sclerosis and spinal cord injury
 PT
 PT
 XX
      Claim 1; Page 41-42; 45pp; English.
  PS
      The present sequence is the protein sequence for human Acute Neuronal
  XX
      Induced Calcium Binding Protein (ANIC-BP). ANIC-BP coding sequence and
  CC
      protein are useful for treating stroke, acute head trauma, multiple
  CC
       sclerosis and spinal cord injury. ANIC-BP coding sequence and protein
  CC
       are also useful as vaccines for inducing an immunological response in a
  CC
  CC
  CC
       mammal.
  XX
                  337 AA;
  SQ
       Sequence
```

```
Length 337;
                            Score 1704; DB 22;
                     100.0%;
 Query Match
                            Pred. No. 1.3e-146;
                     100.0%;
 Best Local Similarity
                                                                0;
                                             Indels
                                                         Gaps
                             Mismatches
                                          0;
 Matches 337; Conservative
                           0;
         1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKB 60
           Qу
          1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKB 60
Db
         61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
           QУ
         61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
Db
        121 AHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
            QУ
        121 AHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
Db
        181 SDAFATFKOLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
            Qу
        181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
Db
        241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPK 300
            QУ
        241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPK 300
Db
         301 LIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
            QУ
         301 LIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
Db
                                                         81.0
 RESULT 4
 AAB94139
     AAB94139 standard; Protein; 289 AA.
 ID
 XX
     AAB94139;
 AC
 XX
                (first entry)
     26-JUN-2001
 DT
 XX
     Human protein sequence SEQ ID NO:14408.
 DΕ
     Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 KW
 XX
     Homo sapiens.
 os
 XX
     EP1074617-A2.
 PN
 XX
      07-FEB-2001.
 PD
 XX
     28-JUL-2000; 2000EP-0116126.
 PF
 XX
                  99JP-0248036.
      29-JUL-1999;
  PR
                  99JP-0300253.
      27-AUG-1999;
  PR
      11-JAN-2000; 2000JP-0118776.
  ₽R
      02-MAY-2000; 2000JP-0183767.
  PR
      09-JUN-2000; 2000JP-0241899.
  PR
  XX
      (HELI-) HELIX RES INST.
  PA
  XX
```

```
Yamamoto J;
    Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K,
                         Wakamatsu A, Nagai K, Otsuki T;
PI
    Ishii S, Sugiyama T,
PI
XX
    WPI; 2001-318749/34.
    Primer sets for synthesizing polynucleotides, particularly the 5602
DR
    full-length cDNAs defined in the specification, and for the detection
XX
    and/or diagnosis of the abnormality of the proteins encoded by the
PT
PT
PT
     full-length cDNAs -
PT
     Claim 8; SEQ ID 14408; 2537pp + CD ROM; English.
XX
     The present invention describes primer sets for synthesising 5602
PS
     full-length cDNAs defined in the specification. Where a primer set
XX
     comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC
     to the complementary strand of a polynucleotide which comprises one of
 CC
     the 5602 nucleotide sequences defined in the specification, where the
 CC
     oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC
     of an oligonucleotide comprising a sequence complementary to the
 CC
      complementary strand of a polynucleotide which comprises a 5'-end
 CC
      sequence and an oligonucleotide comprising a sequence complementary to a
 CC
      polynucleotide which comprises a 3'-end sequence, where the
 CC
      oligonucleotide comprises at least 15 nucleotides and the combination of
 CC
      the 5'-end sequence/3'-end sequence is selected from those defined in
 CC
      the specification. The primer sets can be used in antisense therapy and
 CC
      in gene therapy. The primers are useful for synthesising polynucleotides,
 CC
      particularly full-length cDNAs. The primers are also useful for the
  CC
      detection and/or diagnosis of the abnormality of the proteins encoded by
  CC
      the full-length cDNAs. The primers allow obtaining of the full-length
  CC
      cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
  CC
      AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
  CC
      AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
  CC
      represent oligonucleotides, all of which are used in the exemplification
  CC
  CC
  CC
       of the present invention.
  CC
  XX
                 289 AA;
       Sequence
  SQ
                                                      Length 289;
                                  Score 1466; DB 22;
                           86.0%;
                                  Pred. No. 4.6e-125;
    Query Match
                           99.7%;
                                                                   Gaps
                                                               0;
    Best Local Similarity
                                                      Indels
                                 0; Mismatches
    Matches 288; Conservative
             49 MKEILCGTNEKEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQ 108
                1 MKEILCGTNEKEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQ 60
   QУ
            109 IGTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDF 168
   Db
                61 IGTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLVKIILFSNQFRDF 120
   QУ
             169 FKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQS 228
   Db
                 121 FKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQS 180
    QУ
             229 LKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQ 288
    Db
                 181 LKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQ 240
    QУ
```

Db

```
289 PIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
             QУ
         241 PIVEILLKNOPKLIEFLSSFOKERTDDEOFADEKNYLIKQIRDLKKTAP 289
Db
RESULT 5
AAB48970
     AAB48970 standard; Protein; 341 AA.
ID
XX
     AAB48970;
AC
XX
     27-MAR-2001 (first entry)
DT
     Human ANIC-BP (acute neuronal induced calcium-binding protein).
XX
DE
     Human; acute neuronal induced calcium-binding protein; ANIC-BP;
XX
     Mo25 homologue; HymA homologue; drug screening; stroke;
KW
     acute head trauma; multiple sclerosis; spinal cord injury; vaccine;
KW
KW
     cerebroprotective; neuroprotective.
KW
XX
     Homo sapiens.
 os
 XX
     WO200078947-A1.
 PN
 XX
      28-DEC-2000.
 PD
 XX
      14-JUN-2000; 2000WO-EP05457.
 PF
 XX
                    99EP-0112024.
      22-JUN-1999;
 PR
 XX
      (MERE ) MERCK PATENT GMBH.
 PA
      Den Daas I, Fischer V, Seyfried C, Von Melchner L;
 XX
 PI
 XX
      WPI; 2001-102721/11.
 DR
      N-PSDB; AAC91772.
 DR
      Novel acute neuronal induced calcium binding protein, useful for
 XX
      treating acute head trauma, stroke, multiple sclerosis and spinal cord
 PT
 PT
      injury
 PT
 XX
      Claim 2; Page 37; 50pp; English.
 PS
      The invention relates to human acute neuronal induced calcium-binding
 XX
      protein (ANIC-BP) and to nucleic acid encoding it. The invention
  CC
      also relates to expression systems and recombinant host cells comprising
  CC
      ANIC-BP DNA, the recombinant production of ANIC-BP, antibodies specific
  CC
      for ANIC-BP, fusion proteins comprising ANIC-BP and an immunoglobulin
  CC
       Fc region, and methods of screening for modulators of ANIC-BP function.
  CC
       ANIC-BP has homology and structural similarity to HymA and Mo25 proteins.
  CC
       ANIC-BP proteins and nucleotides are useful for treating stroke and
  CC
       acute head trauma, multiple sclerosis and spinal cord injury. ANIC-BP
  CC
       proteins are useful in screening assays, for identifying membrane bound
  CC
       or soluble receptors, and also in vaccines. ANIC-BP nucleotides are
  CC
       useful as diagnostic reagents, as tools for tissue expression studies,
  CC
       for chromosome localisation studies, as genetic vaccines, and in
  CC
```

the generation of transgenic animals. The present sequence represents

CC CC

```
human ANIC-BP.
CC
XX
             341 AA;
    Sequence
SO
                     81.0%; Score 1381; DB 22;
                                              Length 341;
                      81.0%; Pred. No. 3.2e-117;
 Query Match
                                                                2;
 Best Local Similarity
                                                          Gaps
                                              Indels
                           31; Mismatches
  Matches 273; Conservative
          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
            1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
QУ
         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
Db
            61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
Qу
         120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
Db
                121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
 Çу
         130 ASDAFATFKDLLTRHKVLVADFLEQNYDT1FEDYEKLLQSENYVTKRQSLKLLGELILDR 239
 Db
             181 ASDAFATFKOLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
 Qу
         240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
 Db
             241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
 ОV
 Db
          300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTA 336
             QУ
          301 KLIEFLSKFQNDRTEDEQFNDEKTYLVKQIRDLKRPA 337
  Db
  RESULT 6
      AAE10858 standard; Protein; 496 AA.
  AAE10858
  TD
  XX
      AAE10858;
  AC
  XX
      18-DEC-2001 (first entry)
  DT
      Gal4-human ANIC-BP-1 fusion protein.
  XX
      Human; acute neuronal induced calcium binding protein type 1 ligand;
  DE
  XX
       ANIC-BP-1; human disease; stroke; head trauma; multiple sclerosis;
       Parkinson's disease; Alzheimer's disease; spinal cord injury; vaccine;
  KW
   KW
       gene therapy; fusion protein; Gal4 protein.
   KW
   KW
   XX
       Chimeric - Homo sapiens.
   os
       Chimeric - Unidentified.
   OS
   XX
       WO200170771-A2.
   PN
   XX
       27-SEP-2001.
   PD
   ХX
       20-MAR-2001; 2001WO-EP03149.
   PF'
   XX
```

```
21-MAR-2000; 2000EP-0106110.
PR
    (MERE ) MERCK PATENT GMBH.
XX
PA
                           Hock B;
XX
    Den Daas I, Duecker K,
PΙ
XΧ
    WPI; 2001-607519/69.
     Novel acute neuronal induced calcium binding protein type 1 ligand
DR
     polypeptides, useful in the treatment of stroke, head trauma, multiple
XX
     sclerosis, Parkinson's disease, Alzheimer's disease and spinal cord
PT
PT
 PΤ
     injury
 PT
     Disclosure; Page 42-44; 46pp; English.
 XX
     The invention relates to human acute neuronal induced calcium binding
 PS
     protein type 1 (ANIC-BP-1) ligand polypeptides and polynucleotides.
 XX
     Sequences of the invention are useful for treating human diseases
 CC
      including stroke, head trauma, multiple sclerosis, Parkinson's disease,
 CC
      Alzheimer's disease and spinal cord injury. They are also useful as
 CC
      vaccines. ANIC-BP-1 ligands are useful for identifying membrane bound
 CC
      soluble receptors. Polynucleotides of the invention are useful as
 CC
      diagnostic reagents, for chromosome localization studies, and as
 CC
      valuable tools for tissue expression studies. They are also useful in
  CC
      gene therapy. The present sequence is Gal4-human ANIC-BP-1 fusion
  CC
      protein comprising the Gal4 protein and a C-terminally linked human
  CC
  CC
  CC
      ANIC-BP-1 protein.
  CC
  XX
                 496 AA;
       Sequence
  SQ
                                                    Length 496;
                                 Score 1381; DB 22;
                          81.0%;
                                 Pred. No. 5.2e-117;
    Query Match
                          81.0%;
                                                    Indels
    Best Local Similarity
                                  Mismatches
    Matches 273; Conservative
                               31;
              4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
               156 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 215
   QУ
             60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
   Db
                216 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 275
   QУ
             120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
   Db
                    276 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 335
    QУ
             180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
    Db
                 336 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 395
    Qу
             240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
    Db
                 396 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 455
     QУ
     Db
              300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTA 336
                  111111 IÎ :11:11ÎI 111 11:1ÎIII11: 1
     QУ
              456 KLIEFLSKFQNDRTEDEQFNDEKTYLVKQIRDLKRPA 492
```

Db

```
AAE10859
     AAE10859 standard; Protein; 552 AA.
ID
XX
     AAE10859;
AC
XX
     18-DEC-2001 (first entry)
DT
XX
     LexA-human ANIC-BP-1 fusion protein.
DE
XX
     Human; acute neuronal induced calcium binding protein type 1 ligand;
KW
     ANIC-BP-1; human disease; stroke; head trauma; multiple sclerosis;
KW
     Parkinson's disease; Alzheimer's disease; spinal cord injury; vaccine;
KW
     gene therapy; fusion protein; LexA protein.
ΚW
XX
     Chimeric - Homo sapiens.
OS
     Chimeric - Unidentified.
OS
XX
                     Location/Qualifiers
FH
     Key
                      1..202
FT
     Region
                      /note= "LexA protein"
FT
                      203..552
FT
     Region
                      /note= "Human ANIC-BP-1 protein"
FT
XX
     WO200170771-A2.
PΝ
ХX
     27-SEP-2001.
ÐΩ
XX
     20-MAR-2001; 2001WO-EP03149.
PF
XX
     21-MAR-2000; 2000EP-0106110.
PR
XX
      (MERE ) MERCK PATENT GMBH.
PA
XX
                               Hock B;
      Den Daas I, Duecker K,
 PΙ
XX
      WPI; 2001-607519/69.
 DR
 XX
     Novel acute neuronal induced calcium binding protein type 1 ligand
     polypeptides, useful in the treatment of stroke, head trauma, multiple
 PT
 PT
      sclerosis, Parkinson's disease, Alzheimer's disease and spinal cord
 PT
      injury -
 PT
 XX
      Disclosure; Page 44-46; 46pp; English.
 PS
 XX
      The invention relates to human acute neuronal induced calcium binding
 CC
      protein type 1 (ANIC-BP-1) ligand polypeptides and polynucleotides.
 CC
      Sequences of the invention are useful for treating human diseases
 CC
      including stroke, head trauma, multiple sclerosis, Parkinson's disease,
 CC
      Alzheimer's disease and spinal cord injury. They are also useful as
 CC
      vaccines. ANIC-BP-1 ligands are useful for identifying membrane bound
 CC
      soluble receptors. Polynucleotides of the invention are useful as
 CC
      diagnostic reagents, for chromosome localization studies, and as
 CC
      valuable tools for tissue expression studies. They are also useful in
 CC
      gene therapy. The present sequence is LexA-human ANIC-BP-1 fusion
 CC
```

RESULT 7

```
protein comprising the LexA protein and a C-terminally linked human
CC
    ANIC-BP-1 protein.
CC
XX
             552 AA;
    Sequence
SQ
                             Score 1381; DB 22;
                                               Length 552;
                      81.0%;
 Query Match
                      81.0%; Pred. No. 6e-117;
 Best Local Similarity
                                                                  2:
                                                           Gaps
                               Mismatches
                                               Indels
 Matches 273; Conservative
                           31;
          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
            Qy
         212 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 271
Db
          60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
            Qу
         272 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 331
Db
         120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
                QУ
         332 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 391
Db
         180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
             QУ
         392 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 451
 Db
         240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
             QY
         452 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 511
 Db
         300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTA 336
             1111111 11 :11:1111 111 11:1111111: 1
 QУ
          512 KLIEFLSKFQNDRTEDEQFNDEKTYLVKQIRDLKRPA 548
 Do
 RESULT 8
 AAY94248
     AAY94248 standard; protein; 341 AA.
 ID
 XX
 AC
     AAY94248;
 XX
                 (first entry)
      10-AUG-2000
 DT
 XX
      Mouse calcium binding protein MO25.
 DE
      Mouse; calcium binding protein; cancer; inflammation; MO25; CBP;
 XX
      reproductive disorder; autoimmune disorder; developmental disorder;
  KW
  KW
      seizure disorder; immune disorder; infection.
  KW
  XX
      Mus sp.
  OS
  XX
      WO200029580-A1.
  DN
  XX
      25-MAY-2000.
  PD
  XX
                   99WO-US27027.
      12-NOV-1999;
  PF
  XX
                   98US-0190965.
      13-NOV-1998;
  PR
```

```
XX
    (INCY-) INCYTE PHARM INC.
PΑ
XX
                         Corley NC, Gorgone GA;
             Guegler KJ,
    Tang YT,
PΙ
XX
    WPI; 2000-387793/33.
DR
    Human hCBP protein, and the nucleic acid encoding it, useful for e.g.
XX
    diagnosis, prevention and treatment of cancers, immune, developmental
PT
PT
    or reproductive disorders -
PT
XX
    Disclosure; Page 66-67; 72pp; English.
PS
    The present sequence is the mouse calcium binding protein MO25. It
XX
    was used in a sequence alignment to identify human calcium binding
CC
     protein hCBP. The hCBP protein and the gene encoding it are
CC
     useful for the diagnosis and treatment of the following types of
CC
     disorder: cancers (such as adenocarcinomas), reproductive disorders
CC
     (such as infertility, ovulatory defects, endometriosis, disruptions of
CC
     the oestrus and menstrual cycles, polycystic ovary syndrome and ovarian
CC
     hyperstimulation), autoimmune disorders (such as benign prostatic
CC
     hyperplasia and prostatitis), developmental disorders (such as
CC
     Cushing's syndrome, muscular dystrophy and gonadal dysgenesis),
CC
     hereditary neuropathies, seizure disorders, immune disorders (such as
CC
     AIDS, allergies, anaemia, asthma, atherosclerosis, cholecystitis, Crohn's
CC
     disease, diabetes, Graves' disease, multiple sclerosis, psoriasis,
 CC
     rheumatoid arthritis, scleroderma, Sjogren's syndrome and ulcerative
CC
     colitis), and viral, bacterial, fungal, parasitic, protozoal and
 CC
 CC:
     helminthic infections.
 CC
 XX
               341 AA;
     Sequence
 SQ
                                                  Length 341;
                               Score 1376; DB 21;
                        80.8%;
   Query Match
                                Pred. No. 9e-117;
                        80.7%;
   Best Local Similarity
                                  Mismatches
                                                   Indels
                              32;
   Matches 272; Conservative
            4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
              Qy
            1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
 Db
           60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
              QУ
           61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
 IJb
          120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
                  Qv
          121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
 Db
          180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
              Qу
          181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
 Db
          240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
              QΥ.
           241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQT 300
```

300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTA 336

Db

QУ

3

```
RESULT 9
ABG23844
    ABG23844 standard; Protein; 354 AA.
ID
XX
     ABG23844;
AC
XX
     18-FEB-2002 (first entry)
DT
     Novel human diagnostic protein #23835.
XX
DE
     Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX
     food supplement; medical imaging; diagnostic; genetic disorder.
KW
KW
XX
     Homo sapiens.
os
XX
     WO200175067-A2.
PN
XX
      11-OCT-2001.
 PD
 XX
      30-MAR-2001; 2001WO-US08631.
 PF
 XX
      31-MAR-2000; 2000US-0540217.
 PR
      23-AUG-2000; 2000US-0649167.
 PP.
 XX
      (HYSE-) HYSEQ INC.
 PA
 XX
                            Tang YT;
      Drmanac RT, Liu C,
 PI
 XX
      WPI; 2001-639362/73.
 DR
      N-PSDB; AAS88031.
 DR
      New isolated polynucleotide and encoded polypeptides, useful in
 XX
      diagnostics, forensics, gene mapping, identification of mutations
 PT
       responsible for genetic disorders or other traits and to assess
  PT
  PT
       biodiversity
  PT
       Claim 20; SEQ ID No 54203; 103pp; English.
  XX
  PS
       The invention relates to isolated polynucleotide (I) and
  XX
       polypeptide (II) sequences. (I) is useful as hybridisation probes,
  CC
       polymerase chain reaction (PCR) primers, oligomers, and for chromosome
  CC
       and gene mapping, and in recombinant production of (II). The
  CC
       polynucleotides are also used in diagnostics as expressed sequence tags
  CC
       for identifying expressed genes. (I) is useful in gene therapy techniques
  CC
       to restore normal activity of (II) or to treat disease states involving
  CC
        (II). (II) is useful for generating antibodies against it, detecting or
  CC
       quantitating a polypeptide in tissue, as molecular weight markers and as
  CC
        a food supplement. (II) and its binding partners are useful in medical
  CC .
        imaging of sites expressing (II). (I) and (II) are useful for treating
  CC
        disorders involving aberrant protein expression or biological activity.
  CC
        The polypeptide and polynucleotide sequences have applications in
   CC
        diagnostics, forensics, gene mapping, identification of mutations
   CC
        responsible for genetic disorders or other traits to assess biodiversity
   CC
   CC
```

```
and to produce other types of data and products dependent on DNA and
    amino acid sequences. ABG00010-ABG30377 represent novel human
CC
    diagnostic amino acid sequences of the invention.
    Note: The sequence data for this patent did not appear in the printed
CC
    specification, but was obtained in electronic format directly from WIPO
CC
CC
    at ftp.wipo.int/pub/published_pct_sequences.
CC
CC
XX
              354 AA;
    Sequence
                                                Length 354;
SO
                             Score 1354; DB 22;
                       79.5%;
                       79.2%; Pred. No. 9.5e-115;
  Query Match
                                                            Gaps
                                                         4;
                                                Indels
  Best Local Similarity
                                Mismatches
                            33;
  Matches 267; Conservative
           4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
             14 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 73
 QУ
           50 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
 Db
             74 DPQTEAGAQLAQELYNSGLLITLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 133
 Qу
          120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
 Db
                 134 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLGKIILWSEQFYDFFRYVEMSTFDI 193
  QУ
          130 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
  Db
              294 ASDAFATFKGLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 253
  OV
           240 HNFAIMTKYISKPENLKIMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
  CC
              254 HNFTIMTKYISKPVNLKLMMNLLRDKSRNIQFEAFHVFKAFVANPNKTQPILDILLKNQA 313
  QΨ
  ďď
           300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTA 336
               QУ
           314 KLIEFLSKFONDRTEDEOFNDEKTYLVKOIRDLKRPA 350
   Db
   RESULT 10
       AAB20387 standard; Protein; 350 AA.
   AAB20387
   ID
   XX
        AAB20387;
   AC
   XX
                   (first entry)
        11-JUN-2001
        Human acute neuronal induced calcium binding protein ANIC-BP-1B.
   DT
    XX
        Acute neuronal induced calcium binding protein; ANIC-BP-1B;
    DE
        spice variant; human; stroke; head trauma; Parkinson's disease;
    XX
        Alzheimer's disease; multiple sclerosis; spinal cord injury;
    ΚM
        cerebroprotective; antiparkinsonian; nootropic; neuroprotective;
    ΚW
    KW
        therapy; diagnosis; vaccine.
    KW
    KM
    XX
         Homo sapiers.
    OS
    XX
         WO200125423-A1.
     PN
```

```
XX
    12-APR-2001.
PD
XX
    28-SEP-2000; 2000WO-EP09475.
PF
XX
                  99EP-0119113.
    04-OCT-1999;
PR
XX
     (MERE ) MERCK PATENT GMBH.
PA
XX
    Duecker K, Den Daas I;
PΙ
XX
     WPI; 2001-266306/27.
DR
     N-PSDB; AAF30688.
DR
     Novel human acute neuronal induced calcium-binding protein like protein
ХX
РΤ
     splice variant, useful for treating stroke, acute head trauma,
     Parkinson's disease, Alzheimer's disease multiple sclerosis, spinal
PT
PT
     cord injury -
PT
ХX
     Claim 2; Page 44-45; 49pp; English.
PS
     The present sequence is that of a novel human acute neuronal induced
XX
     calcium binding protein-like protein splice variant, ANIC-NP-1B.
CC
     The protein shows homology to other members of the calcium binding
CC
     protein family, including ANIC-BP, a protein discovered by mRNA
CC
     differential display that is upregulated in a rat model of head
 CC
     trauma. ANIC-BP and ANIC-BP-1B differ in their C-terminal portions.
 CC
     The variant protein could serve as a novel drug target. The
 CC
     invention provides ANIC-BP-1B polynucleotides (see AAF30688) and
 CC.
     polypeptides, expression vectors, host cells and antibodies, as
 CC
     well as methods for producing the protein and for treating or
 CC:
     preventing disorders associated with expression of the protein by
 CC
     inhibiting or activating the action of ANIC-BP-1B. Diseases that
 CC
     may be treated include stroke and acute head trauma, Parkinson's
 CC
      disease, Alzheimer's disease, multiple sclerosis and spinal cord
 CC
              The polynucleotides and polypeptides can also be used in
 CC
      diagnostic assays and in vaccines, and to identify agonists and
 CC
      antagonists useful for treating conditions associated with
 CC
 CC
      ANIC-BP-1B imbalance.
 CC
 XX
                350 AA;
      Sequence
 SO
                                 Score 1297.5; DB 22; Length 350;
                          76.1%;
   Query Match
                          76.0%; Pred. No. 1.3e-109;
   Best Local Similarity
                                                              13; Gaps
                               32; Mismatches
                                                     Indels
                                                38;
   Matches 263; Conservative
             4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
               QΥ
             1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
 Db
            60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
               QУ
            61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
  Db
           120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
                   Qy
           121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
```

Db

```
180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
             Qy
         181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
Db
         240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
             QУ
         241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300
Db
         300 KLIEFLSSFQKERTD------DEQFADEKNYLIKQIRDLKKTA 336
Qу
                                               : |||||
                                    : :
             [[]]
         301 KLIEFLSKFONDRTDCMSSSVPTTNSRVDLRVKPRTRGIRDLKRPA 346
Db
RESULT 11
AAM40864
     AAM40864 standard; Protein; 237 AA.
XX
AC
     AAM40864;
XX
     22-OCT-2001 (first entry)
DT
 XX
     Human polypeptide SEQ ID NO 5795.
 DE
     Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 XX
     peripheral nervous system; neuropathy; central nervous system; CNS;
 ΚW
     Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KM
     amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 ΚW
     chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 WX
 KW
      leukaemia.
 KW
 XX
      Homo sapiens.
 OS'
 XX
      WO200153312-A1.
 PN
 XX
      26-JUL-2001.
 PD
 XX
      26-DEC-2000; 2000WO-US34263.
 PF
 XX
      21-JAN-2000; 2000US-0488725.
 PR
      25-APR-2000; 2000US-0552317.
      09-JUL-2000; 2000US-0598042.
 PR
      19-JUL-2000; 2000US-0620312.
  PR
      03-AUG-2000; 2000US-0653450.
  PR
      14-SEP-2000; 2000US-0662191.
  PR
      19-OCT-2000; 2000US-0693036.
  PR
      29-NOV-2000; 2000US-0727344.
  PR
  XX
       (HYSE-) HYSEQ INC.
  PA
      Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB,
                                                                  Wang D;
  XX
                                                          Ren F,
      Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y,
  PΙ
                                                          Zhang J;
  ΡI
       Zhao QA, Zhou P, Goodrich R, Drmanac RT;
  _{\rm PI}
  XX
       WPI, 2001-442253/47.
  DR
       N-PSDB; AA160020.
  DR
  XX
```

```
Novel nucleic acids and polypeptides, useful for treating disorders
PT
    such as central nervous system injuries -
PT
XX
    Example 2; SEQ ID NO 5795; 10078pp; English.
PS
XX
    The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC
    the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC
    immunosuppressant and cytostatic activity. The polynucleotides are useful
CC
    in gene therapy. A composition containing a polypeptide or polynucleotide
CC
    of the invention may be used to treat diseases of the peripheral nervous
CC
    system, such as peripheral nervous injuries, peripheral neuropathy and
CC
    localised neuropathies and central nervous system diseases, such as
CC
    Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC
    lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC
    utilisation of the activities such as: Immune system suppression,
CC
    Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC
    and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC
    assays for receptor activity, arthritis and inflammation, leukaemias and
CC
    C.N.S disorders.
CC
    Note: The sequence data for this patent did not form part of the printed
CC
     specification.
CC
XX
               237 AA;
     Sequence
SQ
                        68.2%; Score 1162; DB 22;
                                                  Length 237;
  Query Match
                        100.0%; Pred. No. 1.6e-97;
  Best Local Similarity
                             0; Mismatches
                                                  Indels
  Matches 227; Conservative
                                               0;
         111 TRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFK 170
             2 TRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFK 61
Dh
         171 YVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLK 230
QУ
             52 YVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLK 121
Db
          231 LLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPI 290
Qу
             122 LLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPI 181
          291 VEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
QУ
             182 VEILLKNOPKLIEFLSSFOKERTDDEOFADEKNYLIKOIRDLKKTAP 228
Db
RESULT 12
ABE60392
     ABB60392 standard; Protein; 339 AA.
 ID
 XX
     ABE60392;
 AC
 ХX
     26-MAR-2002 (first entry)
 DT
 XX
     Drosophila melanogaster polypeptide SEQ ID NO 7968.
 DE
 XX
     Prosophila; developmental biology; cell signalling; insecticide;
 KW
```

pharmaceutical.

I.W

```
XX
os
    Drosophila melanogaster.
XX
PN
    WO200171042-A2.
XX
PD
    27-SEP-2001.
XX
    23-MAR-2001; 2001WO-US09231.
PF
XX
    23-MAR-2000; 2000US-191637P.
PR
    11-JUL-2000; 2000US-0614150.
PR
XX
    (PEKE ) PE CORP NY.
PΑ
XX
                               Myers EW;
                       Li PWD,
PΙ
    Venter JC, Adams M,
XX
    WPI; 2001-656860/75.
DR
    N-PSDB; ABL04495.
DR
XX
    New isolated nucleic acid detection reagent for detecting 1000 or more
PT
    genes from Drosophila and for elucidating cell signalling and cell-cell
PT
PТ
    interactions -
XX
    Disclosure; SEQ ID NO 7968; 21pp + Sequence Listing; English.
PS
XX
    The invention relates to an isolated nucleic acid detection reagent.
CC
    capable of detecting 1000 or more genes from Drosophila. The invention is
CC
    useful in developmental biology and in elucidating cell signalling and
CC
    cell-cell interactions in higher eukaryotes for the development of
CC
    insecticides, therapeutics and pharmaceutical drugs. The invention
CC
    discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC
    sequences (ABL01840-ABL16175) and the encoded proteins
CC
    (ABB57737-ABB72072).
CC
    The sequence data for this patent did not form part of the printed
CC
    specification, but was obtained in electronic format directly from WIPO
CC
    at ftp.wipo.int/pub/published_pct_sequences.
CC
XX
SQ
    Sequence
              339 AA;
                       65.2%; Score 1111; DB 22; Length 339;
  Query Match
                       65.0%; Pred. No. 1.2e-92;
  Best Local Similarity
  Matches 217; Conservative 59; Mismatches
                                            54;
                                                  Indels
                                                              Gaps . 3;
           4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63
Qу
             1 MPLFGKSQKSPVELVKSLKEAINALEAGDRKVEKAQEDVSKNLVSIKNMLYGSSDAEPPA 60
Db
          64 E-AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 122
QУ
             51 DYVVAQLSQELYNSNLLLLIQNLHRIDFEGKKHVALIFNNVLRRQIGTRSPTVEYICTK 120
Db
         123 PHILFMLLKGYE--APQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
Qγ
                          | | | | | : | | | |
         121 PEILFTLMAGYEDAHPEIALNSGTMLRECARYEALAKIMLHSDEFFKFFRYVEVSTFDIA 180
Db
         181 SDAFATFKDLLTRHKVLVADFLEQNYDTIF-EDYEKLLQSENYVTKRQSLKLLGELILDR 239
Qу
```

```
181 SDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLLNSENYVTRRQSLKLLGELLLDR 240
Db
         240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
Qу
             241 HNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVFVANPNKPKPILDILLRNQT 300
Db
         300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK 333
Qу
             301 KLVDFLTNFHTDRSEDEQFNDEKAYLIKQIKELK 334
Db
RESULT 13
AAY94249
     AAY94249 standard; protein; 339 AA.
ID
XX
AC
    AAY94249;
XX
     10-AUG-2000 (first entry)
DΤ
XX
     Drosophila calcium binding protein DMO25.
DΕ
XX
     Drosophila; calcium binding protein; cancer; inflammation; DMO25; CBP;
KW
     reproductive disorder; autoimmune disorder; developmental disorder;
XW
     seizure disorder; immune disorder; infection.
ΚW
XX
     Drosophila melanogaster.
OS
XX
     W0200029580-A1.
PM
XX
PD
     25-MAY-2000.
XX
PF
     12-NOV-1999;
                   99WO-US27027.
XZ
                   98US-0190965.
PR
     13-NOV-1998;
XX
     (INCY-) INCYTE PHARM INC.
PΑ
XX
     Tang YT, Guegler KJ, Corley NC,
PΙ
XX
DR
     WPI; 2000-387793/33.
XX
     Human hCBP protein, and the nucleic acid encoding it, useful for e.g.
PT
     diagnosis, prevention and treatment of cancers, immune, developmental
PT
     or reproductive disorders -
PT
ХX
     Disclosure; Page 67-68; 72pp; English.
PS
XX
     The present sequence is the Drosophila calcium binding protein DMO25. It
CC
     was used in a sequence alignment to identify human calcium binding
CC
     protein hCBP. The hCBP protein and the gene encoding it are
CC
     useful for the diagnosis and treatment of the following types of
CC
     disorder: cancers (such as adenocarcinomas), reproductive disorders
CC
     (such as infertility, ovulatory defects, endometriosis, disruptions of
CC
     the oestrus and menstrual cycles, polycystic ovary syndrome and ovarian
CC
     hyperstimulation), autoimmune disorders (such as benign prostatic
CC
     hyperplasia and prostatitis), developmental disorders (such as
CC
     Cushing's syndrome, muscular dystrophy and gonadal dysgenesis),
CC
```

```
hereditary neuropathies, seizure disorders, immune disorders (such as
CC
    AIDS, allergies, anaemia, asthma, atherosclerosis, cholecystitis, Crohn's
CC
    disease, diabetes, Graves' disease. multiple sclerosis, psoriasis,
CC
    rheumatoid arthritis, scleroderma, Sjogren's syndrome and ulcerative
CC
    colitis), and viral, bacterial, fungal, parasitic, protozoal and
CC
    helminthic infections.
CC
XX
             339 AA;
    Sequence
SO
                      65.1%; Score 1109; DB 21;
                                              Length 339;
 Ouery Match
                             Pred. No. 1.8e-92;
                      65.0%;
 Best Local Similarity
                                                                  3;
                                                           Gaps
                                               Indels
                               Mismatches
 Matches 217; Conservative
                           59;
          4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEÉVSKSLQAMKEILCGTNEKEPPT 63
QУ
            1 MPLFGKSQKSPVELVKSLKEAINALEAGDRKVEKAQEDVSKNLVSIKNMLHGSSDAEPPA 60
Db
         64 E-AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 122
QУ
            61 DYVVAQLSQELYNSNLLLLLIQNLHRIDFEGKKHVALIFNNLLRRQIGTRSPTVEYICTK 120
Db
         123 PHILFMLLKGYE--APQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
Qy
            121 PEILFTLMAGYEDAHPEIALNSGTMLRECARYEALAKIMLHSDEFFKFFRYVEVSTFDIA 180
Db
         181 SDAFATFKDLLTRHKVLVADFLEQNYDTIF-EDYEKLLQSENYVTKRQSLKLLGELILDR 239
            QУ
         131 SDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLLNSENYVTRRQSLKLLGELLLDR 240
Db
         240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
            QУ
         241 HNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVFVANPNKPKPILDILLRNQT
Db
         300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK 333
ΩУ
             301 KLVDFLTNFHTDRSEDEQFNDEKAYLIKQIKELK 334
Db
 RESULT 14
 AAY94250
     AAY94250 standard; protein; 377 AA.
 AC
     AAY94250;
 ХX
     10-AUG-2000 (first entry)
 DT
 XX
     C. elegans yeast-like calcium binding protein.
 DΕ
 XX
     Calcium binding protein; cancer; inflammation; yeast-like CBP; CBP;
 ΚŴ
     reproductive disorder; autoimmune disorder; developmental disorder;
 KW
     seizure disorder; immune disorder; infection.
 KW
 XX
     Caenorhabditis elegans.
 OS
 XX
     WO200029580-A1.
 PN
 XX
     25-MAY-2000.
 PD
```

```
XX
         12-NOV-1999;
                                   99WO-US27027.
PF
XX
                                   98US-0190965.
         13-NOV-1998;
PR
XX
         (INCY-) INCYTE PHARM INC.
PA
XX
                                                 Corley NC, Gorgone GA;
         Tang YT, Guegler KJ,
PΙ
XX
         WPI; 2000-387793/33.
DR
XX
         Human hCBP protein, and the nucleic acid encoding it, useful for e.g.
PT
         diagnosis, prevention and treatment of cancers, immune, developmental
PΤ
         or reproductive disorders -
PT
XX
         Disclosure; Page 68-69; 72pp; English.
PS
XX
         The present sequence is the C. elegans yeast-like CBP. It
CC
         was used in a sequence alignment to identify human calcium binding
CC
         protein hCBP. The hCBP protein and the gene encoding it are
CC
         useful for the diagnosis and treatment of the following types of
CC
         disorder: cancers (such as adenocarcinomas), reproductive disorders
CC
         (such as infertility, ovulatory defects, endometriosis, disruptions of
CC
         the oestrus and menstrual cycles, polycystic ovary syndrome and ovarian
CC
         hyperstimulation), autoimmune disorders (such as benign prostatic
CC
         hyperplasia and prostatitis), developmental disorders (such as
CC
         Cushing's syndrome, muscular dystrophy and gonadal dysgenesis),
CC
         hereditary neuropathies, seizure disorders, immune disorders (such as
CC
         AIDS, allergies, anaemia, asthma, atherosclerosis, cholecystitis, Crohn's
CC
         disease, diabetes, Graves' disease, multiple sclerosis, psoriasis,
CC
         rheumatoid arthritis, scleroderma, Sjogren's syndrome and ulcerative
CC
         colitis), and viral, bacterial, fungal, parasitic, protozoal and
CC
CC
         helminthic infections.
XX
 SO
         Sequence
                             377 AA;
                                                            Score 1063.5; DB 21;
                                                                                                    Length 377;
                                               62.4%;
    Query Match
                                              60.5%; Pred. No. 2.8e-88;
    Best Local Similarity
                                                                                                                  17; Gaps
                                                         53; Mismatches 68; Indels
    Matches 211; Conservative
                      4 MP-LFSKSHKNPAEIVKILKDNLAILEK-----
                                                                                                  ODKKTDKASEEVSKSLQAM 49
 Qу
                          111 111 : 111::
                      1 MPLLFGKSHKSPADVVKTLREVLTILDKLPPPKLDKDGNIQSDKKYDKALDEVSKNVAMI 60
 Db
                    50 KEILCGTNEKEPPTE---AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILR 106
 Qy
                                                          | : | : | : |
                    61 KSFIYGNDSAEPSSEHVVQVAQLAQEVYNANILPMLIKMLPKFEFECKKDVGQIFNNLLR 120
 Db
                   107 RQIGTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFR 166
 QУ
                          121 RQIGTRSPTVEYLGARPEILIQLVQGYSVPDIALTCGLMLRESIRHDHLAKIILYSDVFY 180
 Db
                   167 DFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKR 226
 QУ
                                           ][]::[]]::[]::[]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[:]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[:::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[::]::[:::[::]::[::]::[::]::[::]::[:::[::]::[::]::[::]::[:::[::]::[::]::[:::[::]::[::]::[:::[::]::[::]::[::]::[::]::[:::[::]::[::]::[::]::[::]::[::]::[:::[::]::[::]::[:::[::]::[::]::[:::[::]::[::]::[:::[::]::[::]::[:::[::]::[:::[::]::[:::[::]::[:::[:::]::[:::[:::]::[:::[:::[:::]::[:::[:::[:::]::[:::[:::[:::]::[:::[:::[:::[:::]::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::[:::
                   181 TFFLYVQSEVFDISSDAFSTFKELTTRHKAIIAEFLDSNYDTFFAQYQNLLNSKNYVTRR 240
 מנו
                   227 QSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHK 286
```

СV

150

```
241 QSLKLLGELLLDRHNFNTMTKYISNPDNLRLMMELLRDKSRNIQYEAFHVFKVFVANPNK 300
Db
         287 TQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKT 335
Qу
               301 PKPISDILNRNREKLVEFLSEFHNDRTDDEQFNDEKAYLIKQIQEMKSS 349
Db
RESULT 15
AAG45273
     AAG45273 standard; Protein; 343 AA.
ID
XX
     AAG45273;
AC
XX
                 (first entry)
DT
     18-OCT-2000
XX
    Arabidopsis thaliana protein fragment SEQ ID NO: 56816.
DE
XX
     Protein identification; signal transduction pathway; metabolic pathway;
KW
     hybridisation assay; genetic mapping; gene expression control; promoter;
KW
     termination sequence.
ΚW
XX
     Arabidopsis thaliana.
OS
XX
PN
     EP1033405-A2.
XX
     06-SEP-2000.
PD
XX
     25-FEB-2000; 2000EP-0301439.
PF
XX
                   99US-0121825.
PR
     25-FEB-1999;
                   99US-0123180.
PR
     ·05-MAR-1999;
                   99US-0123548.
     09-MAR-1999;
PR
     23-MAR-1999;
                   99US-0125788.
PR
                   99US-0126264.
     25-MAR-1999;
₽R
                   99US-0126785.
     29-MAR-1999;
PR
                   99US-0127462.
₽R
     01-APR-1999;
     06-APR-1999;
                   99US-0128234.
PR
PR
     08-APR-1999;
                    99US-0128714.
     16-APR-1999;
                   99US-0129845.
PR
                    99US-0130077.
     19-APR-1999;
PR
                    99US-0130449.
PR
     21-APR-1999;
PR
     23-APR-1999;
                    99US-0130510.
     23-APR-1999;
                    99US-0130891.
₽R
                    99US-0131449.
PR
     28-APR-1999;
                    99US-0132048.
ΡŖ
     30-APR-1999;
PR
     30-APR-1999;
                    99US-0132407.
                    99US-0132484.
PR
     04-MAY-1999;
                    99US-0132485.
PR
     05-MAY-1999;
                    99US-0132486.
PR
     06-MAY-1999;
                    99US-0132487.
PR
     06-MAY-1999;
                    99US-0132863.
PR
     07-MAY-1999;
     11-MAY-1999;
                    99US-0134256.
PR
                    99US-0134218.
PR
     14-MAY-1999;
                    99US-0134219.
PR
     14-MAY-1999;
PR
     14-MAY-1999;
                    99US-0134221.
                    99US-0134370.
PR
     14-MAY-1999;
```

```
99US-0134768.
     18-MAY-1999;
₽R
                     99US-0134941.
     19-MAY-1999;
PR
                     99US-0135124.
PR
     20-MAY-1999;
     21-MAY-1999;
                     99US-0135353.
PR
                      99US-0135629.
     24-MAY-1999;
PR
                      99US-0136021.
     25-MAY-1999;
PR
                     99US-0136392.
PR
     27-MAY-1999;
                      99US-0136782.
     28-MAY-1999;
PR
                      99US-0137222.
     01-JUN-1999;
PR
                     99US-0137528.
PR
     03-JUN-1999;
                      99US-0137502.
PR
     04-JUN-1999;
                      99US-0137724.
     07-JUN-1999;
PR
                      99US-0138094.
     08-JUN-1999;
PR
                      99US-0138540.
     10-JUN-1999;
PR
                      99US-0138847.
     10-JUN-1999;
PR
                      99US-0139119.
     14-JUN-1999;
PR
     16-JUN-1999;
                      99US-0139452.
PR
                      99US-0139453.
     16-JUN-1999;
PR
     17-JUN-1999;
                      99US-0139492.
PR
                      99US-0139454.
PR
     18-JUN-1999;
                      99US-0139455.
     18-JUN-1999;
PR
     18-JUN-1999;
                      99US-0139456.
PR
                      99US-0139457.
     18-JUN-1999;
PR
                      99US-0139458.
     18-JUN-1999;
PR
                      99US-0139459.
PR
     18-JUN-1999;
                      99US-0139460.
     18-JUN-1999;
PR
                      99US-0139461.
PR
      18-JUN-1999;
                      99US-0139462.
PR
     18-JUN-1999;
                      99US-0139463.
PR
      18-JUN-1999;
      18-JUN-1999;
                      99US-0139750.
₽R
PR
      18-JUN-1999;
                      99US-0139763.
PR
      21-JUN-1999;
                      99US-0139817.
PR
      22-JUN-1999;
                      99US-0139899.
ΡR
      23-JUN-1999;
                      99US-0140353.
PR
      23-JUN-1999;
                      99US-0140354.
                      99US-0140695.
PR
      24-JUN-1999;
                      99US-0140823.
PR
      28-JUN-1999;
                      99US-0140991.
PR
      29-JUN-1999;
                      99US-0141287.
PR
      30-JUN-1999;
      01-JUL-1999;
                      99US-0141842.
PR
                      99US-0142154.
      01-JUL-1999;
PR
      02-JUL-1999;
                      99US-0142055.
PR
                      99US-0142390.
      06-JUL-1999;
PR
                      99US-0142803.
      08-JUL-1999;
PR
      09-JUL-1999;
                      99US-0142920.
PR
                      99US-0142977.
      12-JUL-1999;
PR
                      99US-0143542.
      13-JUL-1999;
PR
                      99US-0143624.
PR
      14-JUL-1999;
      15-JUL-1999;
                      99US-0144005.
PR
                      99US-0144085.
      16-JUL-1999;
PR
                      99US-0144086.
      16-JUL-1999;
PR
                      99US-0144325.
PR
      19-JUL-1999;
                      99US-0144331.
      19-JUL-1999;
PR
                      99US-0144332.
PR
      19-JUL-1999;
                      99US-0144333.
      19-JUL-1999;
PR
                      99US-0144334.
PR
      19-JUL-1999;
                      99US-0144335.
ĽR
      19-JUL-1999;
```

ů

٧,

; - -

```
99US-0144352.
     20-JUL-1999;
PR
                      99US-0144632.
     20-JUL-1999;
PR
     20-JUL-1999;
                      99US-0144884.
PR
                      99US-0144814.
     21-JUL-1999;
PR
                      99US-0145086.
     21-JUL-1999;
PR
                      99US-0145088.
     21-JUL-1999;
PR
                      99US-0145085.
     22-JUL-1999;
PR
                      99US-0145087.
     22-JUL-1999;
PR
                      99US-0145089.
     22-JUL-1999;
PR
                      99US-0145192.
     22-JUL-1999;
PR
     23-JUL-1999;
                      99US-0145145.
PR
                      99US-0145218.
     23-JUL-1999;
PR
                      99US-0145224.
     23-JUL-1999;
PR
                      99US-0145276.
     26-JUL-1999;
PR
                      99US-0145913.
PR
     27-JUL-1999;
                      99US-0145918.
PR
     27-JUL-1999;
     27-JUL-1999;
                      99US-0145919.
PR
                      99US-0145951.
PR
     28-JUL-1999;
                      99US-0146386.
PR
      02-AUG-1999;
                      99US-0146388.
      02-AUG-1999;
PR
      02-AUG-1999;
                      99US-0146389.
PR
                      99US-0147038.
PR
      03-AUG-1999;
                      99US-0147204.
      04-AUG-1999;
PR
                      99US-0147302.
      04-AUG-1999;
PR
                      99US-0147192.
PR
      05-AUG-1999;
                      99US-0147260.
      05-AUG-1999;
PR
                      99US-0147303.
      06-AUG-1999;
PR
                      99US-0147416.
PR
      06-AUG-1999;
                      99US-0147493.
      09-AUG-1999;
PR
                      99US-0147935.
      09-AUG-1999;
₽R
                      99US-0148171.
PR
      10-AUG-1999;
                      99US-0148319.
PR
      11-AUG-1999;
                      99US-0148341.
      12-AUG-1999;
PR
                      99US-0148565.
      13-AUG-1999;
PR
      13-AUG-1999;
                      99US-0148684.
PR
                      99US-0149368.
      16-AUG-1999;
PR
                      99US-0149175.
PR
      17-AUG-1999;
      18-AUG-1999;
                      99US-0149426.
PR
PR
      20-AUG-1999;
                      99US-0149722.
                      99US-0149723.
      20-AUG-1999;
PR
                      99US-0149929.
      20-AUG-1999;
PR
                      99US-0149902.
      23-AUG-1999;
PR
                      99US-0149930.
PR
      23-AUG-1999;
      25-AUG-1999;
                      99US-0150566.
PR
                      99US-0150884.
PR
      26-AUG-1999;
                      99US-0151065.
PR
      27-AUG-1999;
                      99US-0151066.
PR
      27-AUG-1999;
                      99US-0151080.
PR
      27-AUG-1999;
                      99US-0151303.
      30-AUG-1999;
₽R
                      99US-0151438.
PR
      31-AUG-1999;
                      99US-0151930.
PR
      01-SEP-1999;
                      99US-0152363.
      07-SEP-1999;
PR
                       99US-0153070.
      10-SEP-1999;
 PR
      13-SEP-1999;
                       99US-0153758.
 PR
                       99US-0154018.
\overline{D}S
      15-SEP-1999;
                       99US-0154039.
 PR
      16-SEP-1999;
                       99US-0154779.
      20-SEP-1999;
 PR
```

.

3. .

```
99US-0155139.
    22-SEP-1999;
PR
                    99US-0155486.
PR
    23-SEP-1999;
                    99US-0155659.
PR
     24-SEP-1999;
                    99US-0156458.
    28-SEP-1999;
PR
                    99US-0156596.
     29-SEP-1999;
PR
                    99US-0157117.
     04-OCT-1999;
PR
                    99US-0157753.
     05-OCT-1999;
PR
                    99US-0157865.
     06-OCT-1999;
PR
                    99US-0158029.
     07-OCT-1999;
PR
                    99US-0158232.
     08-OCT-1999;
PR
                    99US-0158369.
PR
     12-OCT-1999;
                    99US-0159293.
PR
     13-OCT-1999;
                    99US-0159294.
     13-OCT-1999;
PR
                    99US-0159295.
PR
     13-OCT-1999;
                    99US-0159329.
PR
     14-OCT-1999;
                    99US-0159330.
     14-OCT-1999;
PR
                    99US-0159331.
     14-OCT-1999;
PR
     14-OCT-1999;
                    99US-0159637.
PR
                    99US-0159638.
     14-OCT-1999;
PR
                    99US-0159584.
     18-OCT-1999;
PR
                    99US-0160741.
PR
     21-OCT-1999;
                    99US-0160767.
     21-OCT-1999;
PR
                    99US-0160768.
     21-OCT-1999;
PR
                    99US-0160770.
     21-OCT-1999;
PR
                    99US-0160814.
     21-OCT-1999;
PR
                    99US-0160815.
     21-OCT-1999;
PR
                    99US-0160980.
     22-OCT-1999;
\overline{D}\overline{B}
                    99US-0160981.
     22-OCT-1999;
PR
                    99US-0160989.
     22-OCT-1999;
PR
                     99US-0161404.
PR
     25-OCT-1999;
     25-OCT-1999;
                     99US-0161405.
PR
     25-OCT-1999;
                     99US-0161406.
PR
                     99US-0161359.
     26-OCT-1999;
PR
                     99US-0161360.
     26-OCT-1999;
DB
                     99US-0161361.
     26-OCT-1999;
\mathbb{P}\mathbb{R}
                     99US-0161920.
     28-OCT-1999;
PR
                     99US-0161992.
     28-OCT-1999;
PŔ
                     99US-0161993.
     28-OCT-1999;
\mathbb{P}\mathbb{R}
                     99US-0162142.
\mathbb{PR}
     29-OCT-1999;
                                   Score 716.5; DB 21; Length 343;
                           42.0%;
  Query Match
                                   Pred. No. 9.3e-57;
                           42.9%;
  Best Local Similarity
                                                                               3.;
  Matches 144; Conservative 78; Mismatches 105;
                                                                       Gaps
                                                        Indels
             6 LFSKSHKNPAEIVKILKDNLAILEK-----QDKKTDKASEEVSKSLQAMKEILCGTNE 58
 Qу
                                               : ||:||: :| | ::
             4 LFKSKPRTPADIVRQTRDLLLYADRSNSFPDLRESKREEKMVELSKSIRDLKLILYGNSE 63
Db
            59 KEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEY 118
 QУ
                : |
            64 AEPVAEACAQLTQEFFKADTLRRLLTSLPNLNLEARKDATQVVANLQRQQVNSRLIAADY 123
 Db
           119 ISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFD 178
 QУ
                                 : :: :: |: |: |
           124 LESNIDLMDFLVDGFENTDMALHYGTMFRECIRHQIVAKYVLDSEHVKKFFYYIQLPNFD 183
 Db
           179 IASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDY-EKLLQSENYVTKRQSLKLLGELIL 237
 Qу
```

Db	184	:      :  :
Qy		DRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKN 297
Db	244	DRSNSAVMTKYVSSMDNLRILMNLLRESSKTIQIEAFHVFKLFVANQNKPSDIANILVAN 303
Qy	298	QPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK 333 :   :  : ::: :  :  : :
Db	304	RNKLLRLLADIKPDK-EDERFDADKAQVVREIANLK 338

Search completed: January 7, 2004, 16:47:07 Job time: 55 secs

# GenCore version 5.1.6 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 7, 2004, 16:44:17; Search time 21 Seconds

(without alignments)

678.989 Million cell updates/sec

Title: US-10-088-872-2

Perfect score: 1704

Sequence: 1 MKKMPLFSKSHKNPAEIVKI......FADEKNYLIKQIRDLKKTAP 337

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents AA:\*

1: /cgn2 6/ptodata/1/iaa/5A COMB.pep:\*

2: /cgn2 6/ptodata/1/iaa/5B COMB.pep:\*

3: /cgn2 6/ptodata/1/iaa/6A COMB.pep:\*

4: /cgn2 6/ptodata/1/iaa/6B COMB.pep:\*

5: /cgn2 6/ptodata/1/iaa/PCTUS\_COMB.pep:\*

6: /cgn2 6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

		용				
Result		Query				
No.	Score	Match	Length	DB	ID	Description
1	1704	100.0	337	3	US-09-190-965-1	Sequence 1, Appli
2	1704	100.0	337	4	US-09-470-253-1	Sequence 1, Appli
3	1376	80.8	341	3	US-09-190-965-3	Sequence 3, Appli
4	1376	80.8	341	4	US-09-470-253-3	Sequence 3, Appli
5	1109	65.1	339	3	US-09-190-965-4	Sequence 4, Appli
6	1109	65.1	339	4	US-09-470-253-4	Sequence 4, Appli
7	1063.5	62.4	377	3	US-09-190-965-5	Sequence 5, Appli
8	1063.5	62.4	377	4	US-09-470-253-5	Sequence 5, Appli
9	128.5	7.5	3878	4	US-09-914-259-11	Sequence 11, Appl
10	113.5	6.7	1279	4	US-09-724-517-2	Sequence 2, Appli
11	113.5	6.7	1279	4	US-09-641-807A-2	Sequence 2, Appli

12	113.5	6.7	1279	4	US-09-723-096-2	Sequence 2,	Appli
13	113	6.6	2184	4	US-09-417-485D-6	Sequence 6,	Appli
14	105	6.2	586	2	US-08-630-822A-70	Sequence 70	, Appl
15	105	6.2	586	2	US-09-005-069-70	Sequence 70	, Appl
16	105	6.2	586	4	US-09-171-156A-30	Sequence 30	, Appl
17	105	6.2	586	4	US-09-004-730A-30	Sequence 30	, Appl
18	105	6.2	586	4	US-08-981-799A-30	Sequence 30	, Appl
19	103.5	6.1	245	4	US-09-399-913-4	Sequence 4,	Appli
20	103.5	6.1	245	4	US-09-298-731-4	Sequence 4,	Appli
21	103	6.0	387	4	US-09-328-352-5367	Sequence 53	67, Ap
22	103	6.0	2662	4	US-09-595-684B-31	Sequence 31	
23	102.5	6.0	975	4	US-09-914-259-19	Sequence 19	, Appl
24	102.5	6.0	1098	3	US-08-923-992A-8	Sequence 8,	Appli
25	102.5	6.0	1164	3	US-08-923-992A-10	Sequence 10	, Appl
26	102.5	6.0	1388	4	US-09-572-191-2	Sequence 2,	Appli
27	102.5	6.0	1388	4	US-09-723-262-2	Sequence 2,	Appli
28	102.5	6.0	1388	4	US-09-723-219-2	Sequence 2,	Appli
29	102	6.0	474	3	US-08-387-117-6	Sequence 6,	Appli
30	102	6.0	1128	3	US-08-923-992A-6	Sequence 6,	Appli
31	101	5.9	1147	3	US-08-470-260-5	Sequence 5,	Appli
32	101	5.9	1147	3	US-08-471-491-5	Sequence 5,	Appli
33	101	5.9	1147	3	US-08-466-662-5	Sequence 5,	Appli
34	101	5.9	3289	2	US-08-477-451-2	Sequence 2,	Appli
35	99.5	5.8	1164	3	US-08-923-992A-2	Sequence 2,	Appli
36	99	5.8	323	4	US-09-134-001C-3133	Sequence 31	33, Ap
37	98	5.8	1048	3	US-09-356-952-5	Sequence 5,	Appli
38	97	5.7	2482	1	US-08-328-254-6	Sequence 6,	Appli
39	97	5.7	3248	1	US-08-353-700-1	Sequence 1,	Appli
40	97	5.7	3248	5	PCT-US95-16216-1	Sequence 1,	Appli
41	96	5.6	1183	4	US-09-107-532A-6680	Sequence 668	30, Ap
42	95.5	5.6	967	4	US-09-914-259-21	Sequence 21,	. Appl
43	95.5	5.6	1027	4	US-09-914-259-27	Sequence 27	, Appl
44	95	5.6	564	4	US-09-198-452A-601	Sequence 60:	l, App
45	95	5.6	956	4	US-09-914-259-17	Sequence 17,	Appl
						*	

### ALIGNMENTS

```
RESULT 1
US-09-190-965-1
; Sequence 1, Application US/09190965
; Patent No. 6071721
; GENERAL INFORMATION:
 APPLICANT: Tang, Y. Tom
  APPLICANT: Guegler, Karl J.
 APPLICANT: Corley, Neil C.
; APPLICANT: Gorgone, Gina A.
; TITLE OF INVENTION: CALCIUM BINDING PROTEIN
; FILE REFERENCE: PF-0635 US
; CURRENT APPLICATION NUMBER: US/09/190,965
; CURRENT FILING DATE: 1998-11-13
; NUMBER OF SEQ ID NOS: 5
 SOFTWARE: PERL Program
; SEQ ID NO 1
   LENGTH: 337
   TYPE: PRT
```

```
ORGANISM: Homo sapiens
   FEATURE: -
   OTHER INFORMATION: 3734805
US-09-190-965-1
 Query Match 100.0%; Score 1704; DB 3; Length 337; Best Local Similarity 100.0%; Pred. No. 1.6e-161;
 Matches 337; Conservative
                          0; Mismatches
                                           0; Indels
                                                       0;
                                                                 0;
                                                          Gaps
Qу
          1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKE 60
           Db
          1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKE 60
         61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
Qу
            Db
         61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
        121 AHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
Qy
           Db
        121 AHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNOFRDFFKYVELSTFDIA 180
Qy
        181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
           Db
        181 SDAFATFKDLLTRHKVLVADFLEONYDTIFEDYEKLLOSENYVTKROSLKLLGELILDRH 240
        241 NFAIMTKYISKPENLKLMMNLLRDKSPNIOFEAFHVFKVFVASPHKTOPIVEILLKNOPK 300
Qу
           241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPK 300
Db
        301 LIEFLSSFOKERTDDEOFADEKNYLIKOIRDLKKTAP 337
Qy
           Db
        301 LIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
RESULT 2
US-09-470-253-1
; Sequence 1, Application US/09470253
; Patent No. 6365371
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
  APPLICANT: Gorgone, Gina A.
  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
  FILE REFERENCE: PF-0635 US
  CURRENT APPLICATION NUMBER: US/09/470,253
  CURRENT FILING DATE: 1999-12-22
  PRIOR APPLICATION NUMBER: 09/190,965
  PRIOR FILING DATE: 1998-11-13
  NUMBER OF SEQ ID NOS: 5
  SOFTWARE: PERL Program
; SEO ID NO 1
  LENGTH: 337
   TYPE: PRT
   ORGANISM: Homo sapiens
   FEATURE: -
```

OTHER INFORMATION: 3734805

```
Query Match
                      100.0%; Score 1704; DB 4;
                                              Length 337;
  Best Local Similarity
                      100.0%; Pred. No. 1.6e-161;
  Matches 337; Conservative
                           0; Mismatches
                                          0;
                                              Indels
                                                      0;
                                                                0;
Qy
          1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKE 60
            Db
          1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKE 60
Qy
         61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
            Db
         61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
Qу
        121 AHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
            Db
        121 AHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
        181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
Qу
            181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
Db
        241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPK 300
Qy
            Db
        241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPK 300
Qу
        301 LIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
            Db
        301 LIEFLSSFOKERTDDEOFADEKNYLIKOIRDLKKTAP 337
RESULT 3
US-09-190-965-3
; Sequence 3, Application US/09190965
; Patent No. 6071721
; GENERAL INFORMATION:
  APPLICANT: Tang, Y. Tom
  APPLICANT: Guegler, Karl J.
  APPLICANT: Corley, Neil C.
  APPLICANT: Gorgone, Gina A.
  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
  FILE REFERENCE: PF-0635 US
  CURRENT APPLICATION NUMBER: US/09/190.965
  CURRENT FILING DATE: 1998-11-13
  NUMBER OF SEQ ID NOS: 5
  SOFTWARE: PERL Program
 SEO ID NO 3
   LENGTH: 341
   TYPE: PRT
   ORGANISM: Mus sp.
   FEATURE: -
   OTHER INFORMATION: q262934
US-09-190-965-3
 Query Match
                     80.8%; Score 1376; DB 3; Length 341;
 Best Local Similarity
                     80.7%; Pred. No. 7.3e-129;
 Matches 272; Conservative 32; Mismatches
                                         29;
                                             Indels
                                                         Gaps
                                                                2;
```

```
Qу
          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
           Db
          1 MPFPFGKSHKSPADIVKNLKESMAVLEKODISDKKAEKATEEVSKNLVAMKETLYGTNEK 60
Qу
         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
           Dh
         61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAOIFNNILRROIGTRTPTVEYI 120
        120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
Qу
               121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
Db
        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
Qу
           Dh
        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKROSLKLLGELLLDR 240
        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNOP 299
Qу
           Db
        241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNOT 300
        300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTA 336
Qy
           301 KLIEFLSKFQNDRTEDEQFNDEKTYLVKQIRNLKRAA 337
Db
RESULT 4
US-09-470-253-3
; Sequence 3, Application US/09470253
; Patent No. 6365371
; GENERAL INFORMATION:
  APPLICANT: Tang, Y. Tom
  APPLICANT: Guegler, Karl J.
  APPLICANT: Corley, Neil C.
  APPLICANT: Gorgone, Gina A.
  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
  FILE REFERENCE: PF-0635 US
  CURRENT APPLICATION NUMBER: US/09/470,253
  CURRENT FILING DATE: 1999-12-22
  PRIOR APPLICATION NUMBER: 09/190,965
  PRIOR FILING DATE: 1998-11-13
  NUMBER OF SEQ ID NOS: 5
  SOFTWARE: PERL Program
; SEO ID NO 3
   LENGTH: 341
   TYPE: PRT
   ORGANISM: Mus sp.
   FEATURE: -
   OTHER INFORMATION: q262934
US-09-470-253-3
                     80.8%; Score 1376; DB 4; Length 341;
 Query Match
 Best Local Similarity
                    80.7%; Pred. No. 7.3e-129;
 Matches 272; Conservative 32; Mismatches
                                        29; Indels
                                                       Gaps
                                                              2;
Qy
         4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
```

```
1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
Db
Qy
         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
            Db
         61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
Qу
        120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNOFRDFFKYVELSTFDI 179
               121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
Db
        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKROSLKLLGELILDR 239
Qу
            181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
Db
        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTOPIVEILLKNOP 299
Qу
            241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNOT 300
Db
        300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTA 336
Qу
            111111 11 :11:111 111 11:111:11: 1
Dh
        301 KLIEFLSKFQNDRTEDEQFNDEKTYLVKQIRNLKRAA 337
RESULT 5
US-09-190-965-4
; Sequence 4, Application US/09190965
; Patent No. 6071721
; GENERAL INFORMATION:
  APPLICANT: Tang, Y. Tom
  APPLICANT: Guegler, Karl J.
  APPLICANT: Corley, Neil C.
  APPLICANT: Gorgone, Gina A.
  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
  FILE REFERENCE: PF-0635 US
  CURRENT APPLICATION NUMBER: US/09/190,965
  CURRENT FILING DATE: 1998-11-13
  NUMBER OF SEQ ID NOS: 5
  SOFTWARE: PERL Program
; SEQ ID NO 4
   LENGTH: 339
   TYPE: PRT
   ORGANISM: Drosophila melanogaster
   FEATURE: -
   OTHER INFORMATION: g1794137
US-09-190-965-4
 Query Match
                     65.1%; Score 1109; DB 3; Length 339;
 Best Local Similarity
                    65.0%; Pred. No. 2.8e-102;
 Matches 217; Conservative 59; Mismatches
                                         54; Indels
                                                      4; Gaps
                                                                3;
QУ
          4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63
           Db
          1 MPLFGKSQKSPVELVKSLKEAINALEAGDRKVEKAQEDVSKNLVSIKNMLHGSSDAEPPA 60
         64 E-AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 122
Qy
           Db
         61 DYVVAQLSQELYNSNLLLLLIQNLHRIDFEGKKHVALIFNNLLRRQIGTRSPTVEYICTK 120
```

```
QУ
        123 PHILFMLLKGYE--APQIALRCGIMLRECIRHEPLAKIILFSNOFRDFFKYVELSTFDIA 180
           Db
        121 PEILFTLMAGYEDAHPEIALNSGTMLRECARYEALAKIMLHSDEFFKFFRYVEVSTFDIA 180
        181 SDAFATFKDLLTRHKVLVADFLEQNYDTIF-EDYEKLLQSENYVTKROSLKLLGELILDR 239
Qу
           181 SDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLLNSENYVTRRQSLKLLGELLLDR 240
Db
        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
Qy
           Db
        241 HNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVFVANPNKPKPILDILLRNOT 300
Qу
        300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK 333
           Db
        301 KLVDFLTNFHTDRSEDEQFNDEKAYLIKQIKELK 334
RESULT 6
US-09-470-253-4
; Sequence 4, Application US/09470253
; Patent No. 6365371
; GENERAL INFORMATION:
  APPLICANT: Tang, Y. Tom
  APPLICANT: Guegler, Karl J.
  APPLICANT: Corley, Neil C.
  APPLICANT: Gorgone, Gina A.
  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
  FILE REFERENCE: PF-0635 US
  CURRENT APPLICATION NUMBER: US/09/470,253
  CURRENT FILING DATE: 1999-12-22
  PRIOR APPLICATION NUMBER: 09/190.965
  PRIOR FILING DATE: 1998-11-13
  NUMBER OF SEQ ID NOS: 5
  SOFTWARE: PERL Program
; SEQ ID NO 4
   LENGTH: 339
   TYPE: PRT
   ORGANISM: Drosophila melanogaster
   FEATURE: -
   OTHER INFORMATION: q1794137
US-09-470-253-4
 Query Match
                     65.1%; Score 1109; DB 4; Length 339;
 Best Local Similarity 65.0%; Pred. No. 2.8e-102;
 Matches 217; Conservative 59; Mismatches
                                        54;
                                           Indels
                                                     4; Gaps
                                                              3;
Qу
         4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63
           Db
         1 MPLFGKSQKSPVELVKSLKEAINALEAGDRKVEKAQEDVSKNLVSIKNMLHGSSDAEPPA 60
         64 E-AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 122
Qy
           Db
         61 DYVVAQLSQELYNSNLLLLIQNLHRIDFEGKKHVALIFNNLLRRQIGTRSPTVEYICTK 120
Qy
        123 PHILFMLLKGYE--APQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
```

```
Db .
        121 PEILFTLMAGYEDAHPEIALNSGTMLRECARYEALAKIMLHSDEFFKFFRYVEVSTFDIA 180
Qу
        181 SDAFATFKDLLTRHKVLVADFLEQNYDTIF-EDYEKLLOSENYVTKROSLKLLGELILDR 239
            Db
        181 SDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLLNSENYVTRROSLKLLGELLLDR 240
Qy
        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTOPIVEILLKNQP 299
            111 : [1:11:11:11] [1:1:1] [1:1:1] [1:1:1] [1:1:1] [1:1:1]
Db
        241 HNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVFVANPNKPKPILDILLRNQT 300
Qу
        300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK 333
            Db
        301 KLVDFLTNFHTDRSEDEQFNDEKAYLIKQIKELK 334
RESULT 7
US-09-190-965-5
; Sequence 5, Application US/09190965
; Patent No. 6071721
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Guegler, Karl J.
 APPLICANT: Corley, Neil C.
 APPLICANT: Gorgone, Gina A.
  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
  FILE REFERENCE: PF-0635 US
  CURRENT APPLICATION NUMBER: US/09/190,965
  CURRENT FILING DATE: 1998-11-13
  NUMBER OF SEQ ID NOS: 5
 SOFTWARE: PERL Program
; SEQ ID NO 5
  LENGTH: 377
  TYPE: PRT
  ORGANISM: Caenorhabditis elegans
   FEATURE: -
  OTHER INFORMATION: q1255838
US-09-190-965-5
 Query Match
                     62.4%; Score 1063.5; DB 3; Length 377;
 Best Local Similarity 60.5%; Pred. No. 1.1e-97;
 Matches 211; Conservative 53; Mismatches 68; Indels
          4 MP-LFSKSHKNPAEIVKILKDNLAILEK-----QDKKTDKASEEVSKSLQAM 49
Qy
           1 MPLLFGKSHKSPADVVKTLREVLTILDKLPPPKLDKDGNIQSDKKYDKALDEVSKNVAMI 60
Db
         50 KEILCGTNEKEPPTE---AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILR 106
Qу
           Db
         61 KSFIYGNDSAEPSSEHVVQVAQLAQEVYNANILPMLIKMLPKFEFECKKDVGQIFNNLLR 120
QУ
        107 RQIGTRSPTVEYISAHPHILFMLLKGYEAPOIALRCGIMLRECIRHEPLAKIILFSNOFR 166
           Db
        121 RQIGTRSPTVEYLGARPEILIQLVQGYSVPDIALTCGLMLRESIRHDHLAKIILYSDVFY 180
        167 DFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEONYDTIFEDYEKLLOSENYVTKR 226
Qу
            Db
        181 TFFLYVQSEVFDISSDAFSTFKELTTRHKAIIAEFLDSNYDTFFAQYONLLNSKNYVTRR 240
```

```
Qу
        227 QSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIOFEAFHVFKVFVASPHK 286
            241 QSLKLLGELLLDRHNFNTMTKYISNPDNLRLMMELLRDKSRNIQYEAFHVFKVFVANPNK 300
Db
        287 TQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKT 335
Qу
             Db
        301 PKPISDILNRNREKLVEFLSEFHNDRTDDEQFNDEKAYLIKOIOEMKSS 349
RESULT 8
US-09-470-253-5
; Sequence 5, Application US/09470253
; Patent No. 6365371
; GENERAL INFORMATION:
  APPLICANT: Tang, Y. Tom
  APPLICANT: Guegler, Karl J.
  APPLICANT: Corley, Neil C.
  APPLICANT: Gorgone, Gina A.
  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
  FILE REFERENCE: PF-0635 US
  CURRENT APPLICATION NUMBER: US/09/470,253
  CURRENT FILING DATE: 1999-12-22
  PRIOR APPLICATION NUMBER: 09/190,965
  PRIOR FILING DATE: 1998-11-13
  NUMBER OF SEQ ID NOS: 5
  SOFTWARE: PERL Program
 SEQ ID NO 5
   LENGTH: 377
   TYPE: PRT
   ORGANISM: Caenorhabditis elegans
   FEATURE: -
   OTHER INFORMATION: q1255838
US-09-470-253-5
                     62.4%; Score 1063.5; DB 4; Length 377; 60.5%; Pred. No. 1.1e-97;
 Query Match
 Best Local Similarity
 Matches 211; Conservative 53; Mismatches
                                         68; Indels
                                                     17:
                                                         Gaps
                                                                3;
          4 MP-LFSKSHKNPAEIVKILKDNLAILEK-----QDKKTDKASEEVSKSLQAM 49
Qу
           111 111 :1111:: :
          1 MPLLFGKSHKSPADVVKTLREVLTILDKLPPPKLDKDGNIQSDKKYDKALDEVSKNVAMI 60
Db
         50 KEILCGTNEKEPPTE---AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILR 106
Qу
           1:1:1:1
                          Db
         61 KSFIYGNDSAEPSSEHVVQVAQLAQEVYNANILPMLIKMLPKFEFECKKDVGOIFNNLLR 120
        107 RQIGTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNOFR 166
Qу
           Db
        121 RQIGTRSPTVEYLGARPEILIQLVQGYSVPDIALTCGLMLRESIRHDHLAKIILYSDVFY 180
        167 DFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKR 226
Qy
                   11 11:
Db
        181 TFFLYVQSEVFDISSDAFSTFKELTTRHKAIIAEFLDSNYDTFFAQYQNLLNSKNYVTRR 240
Qy
        227 QSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHK 286
```

```
Db
        241 QSLKLLGELLLDRHNFNTMTKYISNPDNLRLMMELLRDKSRNIQYEAFHVFKVFVANPNK 300
QУ
        287 TQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKT 335
            301 PKPISDILNRNREKLVEFLSEFHNDRTDDEQFNDEKAYLIKQIQEMKSS 349
RESULT 9
US-09-914-259-11
; Sequence 11, Application US/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
 APPLICANT: Hyman, Paul
 APPLICANT: Williams, Mark
  TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
  FILE REFERENCE: 8471-010-999
 CURRENT APPLICATION NUMBER: US/09/914,259
; CURRENT FILING DATE: 2000-11-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
  LENGTH: 3878
   TYPE: PRT
   ORGANISM: Homo sapiens
US-09-914-259-11
 Query Match
                    7.5%; Score 128.5; DB 4; Length 3878;
 Best Local Similarity 20.1%; Pred. No. 0.0037;
 Matches 77; Conservative 75; Mismatches 116; Indels 115; Gaps 15;
QУ
         18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77
           Db
        664 IEKLKDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKONOLILE---- 710
        78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI-----GTRSPTVEYISAHPHI 125
QУ
            Db
        711 --ISKLKDLQQSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 766
        126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
Qγ
                               | :: | :| |
        767 LEKOMKEKE-----NDLQEKFAQLEAEN-SILKDEKK 797
Db
        186 TFKDLLTRH-----KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237
Qу
           798 TLEDMLKIHTPVSQEERLIFLDSIKSKSKDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 857
Db
        238 DRHNFAIMTK-----YISKPENLKLMMNLLRD 264
QУ
            1: 1: 1
                                              | | | : | :: | |
        858 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNKQELEYKSKLKALNEELHLQRI 917
Db
Qy
        265 KSPNIOFEA--FHVFKVFVASPHKTOPIVEILLKNOPKLIEFLSSFOKERTD-DEOFAD- 320
              Db
       918 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVVEKDTTELMEKLEVTKREKLELSQRLSDL 974
Qy
        321 -----EKNYLIKQIRDLKK 334
                  1 ::1 :::: 11:
```

```
RESULT 10
US-09-724-517-2
; Sequence 2, Application US/09724517
; Patent No. 6379941
; GENERAL INFORMATION:
  APPLICANT: Beraud, Christophe
  APPLICANT: Freedman, Richard
  TITLE OF INVENTION: No. 6379941el motor proteins and methods for
  TITLE OF INVENTION: their use
  FILE REFERENCE: 1031
  CURRENT APPLICATION NUMBER: US/09/724,517
  CURRENT FILING DATE: 2000-11-27
  PRIOR APPLICATION NUMBER: US/09/641,807
  PRIOR FILING DATE: 2000-08-17
  NUMBER OF SEQ ID NOS: 4
  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
   LENGTH: 1279
   TYPE: PRT
   ORGANISM: Human
   FEATURE:
   NAME/KEY: VARIANT
   LOCATION: (409)...(436)
   OTHER INFORMATION: Xaa = any amino acid
US-09-724-517-2
 Query Match
                      6.7%; Score 113.5; DB 4; Length 1279;
 Best Local Similarity 19.3%; Pred. No. 0.024;
 Matches 87; Conservative 61; Mismatches 137; Indels 165; Gaps 14;
         23 DNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSGLLVTL 82
Qу
            1:1 1::1 1::1 1::1: ::1: ::1
        794 DHLQKLDEQKKWLDEEVEKVLNQRQELEELEADLKKREAIVSKKEALLQE--KSHLENKK 851
Db
Qy
         83 IADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISA----- 121
            : | :: : | |:: |:| :: :: :: :|
Db
        852 LRSSQALNTDSLKISTRL--NLLEQELSEKNVQLQTSTAEEKTKISEOVEVLOKEKDOLO 909
        122 -----HPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFS 162
Qу
                              1:11 | :1 | 1 : 1:
                                                   | | | :: | |
        910 KRRHDVDEKLKNGRVLSPEEEHVLFQLEEGIEALEAAIE---YRNESIQNRQKSLRASFH 966
        163 NQFRDFFKYVE----LSTFDIASDAFATFKDLLT------RHKVLVAD--- 200
QУ
            967 NLSRGEANVLEKLACLSPVEIRTILFRYFNKVVNLREAERKQQLYNEEMKMKVLERDNMV 1026
        201 -----FEDYEKLLQS 219
Qу
                                           1 | | : | : | : | : | :
       1027 RELESALDHLKLQCDRRLTLQQKEHEQKMQLLLHHFKEQDGEGIMETFKTYEDKIQQLEK 1086
Db
        220 ENYVTKRQS-----TKYISK 251
Qу
            : | |: | : | | | |
Db
       1087 DLYFYKKTSRDHKKKLKELVGEAI--RRQLAPSEYQEAGDGVLKPEGGGMLSEELKWASR 1144
```

```
252 PENLKLMMNLLRDKSPNIOFEAFHVFKVFVASPHKTQPIVEILLKNOPKLIEFLSSF--- 308
Qу
                                   ||::|| |:: :
       1145 PESMKLSG---REREMDSS-----ASSLRTQPNPQKLWEDIPELPPIHSSLAPP 1190
Db
Qу
       309 -----OKERTDDEOFADEKNYLIKQIR 330
                  1191 SGHMLGNENKTETDDNQFTKSHSRLSSQIQ 1220
Db
RESULT 11
US-09-641-807A-2
; Sequence 2, Application US/09641807A
; Patent No. 6440731
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
  APPLICANT: Freedman, Richard
  TITLE OF INVENTION: No. 6440731el motor proteins and methods for
  TITLE OF INVENTION: their use
; FILE REFERENCE: 1031
; CURRENT APPLICATION NUMBER: US/09/641,807A
; CURRENT FILING DATE: 2000-08-17
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
  LENGTH: 1279
  TYPE: PRT
  ORGANISM: Human
  FEATURE:
 NAME/KEY: VARIANT
; LOCATION: (409)...(446)
; OTHER INFORMATION: Xaa = any amino acid
US-09-641-807A-2
                     6.7%; Score 113.5; DB 4; Length 1279;
 Query Match
 Best Local Similarity 19.3%; Pred. No. 0.024;
 Matches 87; Conservative 61; Mismatches 137; Indels 165; Gaps 14;
         23 DNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSGLLVTL 82
Qу
           |:| |::| | |: |:| | ::|: |:|
        794 DHLQKLDEQKKWLDEEVEKVLNQRQELEELEADLKKREAIVSKKEALLQE--KSHLENKK 851
Db
         83 IADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISA---- 121
Qу
           : | :: : | |:: |:| ::: :: :|
        852 LRSSOALNTDSLKISTRL--NLLEOELSEKNVOLOTSTAEEKTKISEQVEVLQKEKDQLQ 909
        Qy
                            910 KRRHDVDEKLKNGRVLSPEEEHVLFQLEEGIEALEAAIE---YRNESIQNRQKSLRASFH 966
Db
        163 NQFRDFFKYVE----LSTFDIASDAFATFKDLLT------RHKVLVAD--- 200
Qу
           967 NLSRGEANVLEKLACLSPVEIRTILFRYFNKVVNLREAERKQQLYNEEMKMKVLERDNMV 1026
Db
                             ----FLEQNYDTI----FEDYEKLLQS 219
Qу
                                        1 11::1::1:
Db
       1027 RELESALDHLKLQCDRRLTLQQKEHEQKMQLLLHHFKEQDGEGIMETFKTYEDKIQQLEK 1086
```

```
Qy
        220 ENYVTKRQS-----TKYISK 251
            : | |: | : |: || |
Db
        1087 DLYFYKKTSRDHKKKLKELVGEAI--RRQLAPSEYQEAGDGVLKPEGGGMLSEELKWASR 1144
        252 PENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSF--- 308
Qy
            1145 PESMKLSG---REREMDSS-----ASSLRTQPNPQKLWEDIPELPPIHSSLAPP 1190
Db
Qу
        309 ----QKERTDDEQFADEKNYLIKQIR 330
                   Db
       1191 SGHMLGNENKTETDDNQFTKSHSRLSSQIQ 1220
RESULT 12
US-09-723-096-2
; Sequence 2, Application US/09723096
; Patent No. 6448026
; GENERAL INFORMATION:
; APPLICANT: Beraud, Christophe
; APPLICANT: Freedman, Richard
  TITLE OF INVENTION: No. 6448026el motor proteins and methods for
  TITLE OF INVENTION: their use
; FILE REFERENCE: 1031
  CURRENT APPLICATION NUMBER: US/09/723,096
  CURRENT FILING DATE: 2000-11-27
  PRIOR APPLICATION NUMBER: US/09/641,807
  PRIOR FILING DATE: 2000-08-17
  NUMBER OF SEQ ID NOS: 4
  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
  LENGTH: 1279
   TYPE: PRT
   ORGANISM: Human
   FEATURE:
   NAME/KEY: VARIANT
   LOCATION: (409)...(436)
   OTHER INFORMATION: Xaa = any amino acid
US-09-723-096-2
 Query Match
                     6.7%; Score 113.5; DB 4; Length 1279;
 Best Local Similarity 19.3%; Pred. No. 0.024;
 Matches 87; Conservative 61; Mismatches 137; Indels 165; Gaps 14;
Qу
         23 DNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSGLLVTL 82
            1:1 1::1 1: 1:1 1::1: ::1
        794 DHLQKLDEQKKWLDEEVEKVLNQRQELEELEADLKKREAIVSKKEALLQE--KSHLENKK 851
Db
         83 IADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISA----- 121
Qу
            : | :: : | |:: |:| :: :: ::
Db
        852 LRSSQALNTDSLKISTRL--NLLEQELSEKNVQLQTSTAEEKTKISEQVEVLQKEKDQLQ 909
        122 -----HPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFS 162
Qу
                             Db
        910 KRRHDVDEKLKNGRVLSPEEEHVLFQLEEGIEALEAAIE---YRNESIQNRQKSLRASFH 966
Qу
        163 NQFRDFFKYVE----LSTFDIASDAFATFKDLLT------RHKVLVAD--- 200
           1 | :| :| : | ::
                                                    : ||| |
```

```
967 NLSRGEANVLEKLACLSPVEIRTILFRYFNKVVNLREAERKQQLYNEEMKMKVLERDNMV 1026
Db
Qу
        201 -----FEDYEKLLQS 219
                                          1.11::1::1:
        1027 RELESALDHLKLQCDRRLTLQQKEHEQKMQLLLHHFKEQDGEGIMETFKTYEDKIQQLEK 1086
Db
Qy
        220 ENYVTKRQS------TKYLSELILDRHNFAIM-----TKYISK 251
            : | |: | : |: | | | |
                                                           1: 1:
Db
        1087 DLYFYKKTSRDHKKKLKELVGEAI--RRQLAPSEYQEAGDGVLKPEGGGMLSEELKWASR 1144
        252 PENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSF--- 308
Qу
          1145 PESMKLSG---REREMDSS-----ASSLRTQPNPQKLWEDIPELPPIHSSLAPP 1190
Db
        309 ----QKERTDDEQFADEKNYLIKQIR 330
Qу
                   Db
       1191 SGHMLGNENKTETDDNQFTKSHSRLSSQIQ 1220
RESULT 13
US-09-417-485D-6
; Sequence 6, Application US/09417485D
; Patent No. 6541202
; GENERAL INFORMATION:
; APPLICANT: Long, David M.
; APPLICANT: Metz, Anneke M.
 APPLICANT: Love, Ruschelle A.
  TITLE OF INVENTION: Telomerase Reverse Transcriptase (TERT) Genes
  FILE REFERENCE: 47714-5009-US
; CURRENT APPLICATION NUMBER: US/09/417,485D
; CURRENT FILING DATE: 2002-06-14
; NUMBER OF SEQ ID NOS: 49
 SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
  LENGTH: 2184
   TYPE: PRT
   ORGANISM: Plasmodium falciparum
   FEATURE:
   NAME/KEY: unsure
   LOCATION: (330)..(335)
   OTHER INFORMATION: Xaa at position 330 = Leu or Ile;
   OTHER INFORMATION: Xaa at position 335 = Asp or Gly.
US-09-417-485D-6
 Query Match
                      6.6%; Score 113; DB 4; Length 2184;
 Best Local Similarity 21.9%; Pred. No. 0.057;
 Matches 77; Conservative 58; Mismatches 140; Indels 76; Gaps 17;
          1 MKKMPLFSKSHKNPAEIV--KILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNE 58
Qγ
            Db
        309 LPEIDFFSEDRKEKSSSVGYDXKKKNXSNIKRFHNKINRTKEEKKKKWN--KIIINRNNI 366
Qу
         59 KEPPTEAVAQLAQELYSSGLLVTLIAD------LQLIDFEGKKDVTQIFNN------ 103
            : | : |:: |: |:: |::: |:::|
Db
        367 LQHNT--TNKCKTFLFNKHIIFDKIENNNIPLFIYDLLNYIFKSDQTYFYHNNFIDEYKQ 424
Qу
        104 ILRRQI--GTRSPTVEYI--SAHPHILFMLLK---GYEAPQIALRCGIMLRECIRHEPLA 156
```

```
Db
         425 KICKQIKCSTKKNDISHIITSRKENHLFHVQKLENNYKHPNI-----NKQLRKTKIL 476
QУ
         157 KIILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTR-HKV-----L 197
            1: | |::| | :| | | |:|:|:
         477 KYVY--NYFKEFINNVINTKFGKIYRKFFPRKHILNKIHKIFKIIRLQIIKKYRIINIRM 534
        198 VADFLEQN-YDTIFEDYE-----KLLQSENYVTKR-QSLKLLGELILDRHNFAIMT 246
Qу
               |::| ||| |::|: |:|| |:: |:|| |
Db
        535 NRKFIKQKVYDTFFKNYDFLSFSFKTYKIINFMVYITKKCIPIKLLG----SKHNFKIFL 590
        247 KYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKN 297
Qy
            591 KNVKK----FLLFNYKESFSLNQVMKNIKVKNIFQKKISKYNIKNRILLKN 637
Db
RESULT 14
US-08-630-822A-70
; Sequence 70, Application US/08630822A
; Patent No. 5840695
; GENERAL INFORMATION:
   APPLICANT: FRANK, GLENN R.
    APPLICANT: HUNTER, SHIRLEY WU
   APPLICANT: WALLENFELS, LYNDA
    TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
    TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
    NUMBER OF SEQUENCES: 107
    CORRESPONDENCE ADDRESS:
    ADDRESSEE: Sheridan Ross P.C.
     STREET: 1700 Lincoln Street, Suite 3500
;
    CITY: Denver
     STATE: Colorado
     COUNTRY: U.S.A.
      ZIP: 80203
    COMPUTER READABLE FORM:
     MEDIUM TYPE: Floppy disk
     COMPUTER: IBM PC compatible
     OPERATING SYSTEM: PC-DOS/MS-DOS
     SOFTWARE: PatentIn Release #1.0, Version #1.25
   CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/08/630,822A
    FILING DATE: 11-APR-1996
     CLASSIFICATION: 435
    ATTORNEY/AGENT INFORMATION:
     NAME: CONNELL, GARY J.
     REGISTRATION NUMBER: 32,020
     REFERENCE/DOCKET NUMBER: 2618-17-C3
    TELECOMMUNICATION INFORMATION:
     TELEPHONE: (303) 863-9700
     TELEFAX: (303) 863-0223
  INFORMATION FOR SEQ ID NO: 70:
    SEQUENCE CHARACTERISTICS:
    LENGTH: 586 amino acids
     TYPE: amino acid
    TOPOLOGY: linear
   MOLECULE TYPE: protein
   FEATURE:
```

```
NAME/KEY: Xaa = any amino acid
     LOCATION: 379
US-08-630-822A-70
 Query Match
                     6.2%; Score 105; DB 2; Length 586;
 Best Local Similarity 20.0%; Pred. No. 0.054;
 Matches 77; Conservative 54; Mismatches 136; Indels 118; Gaps 15;
Qy
        22 KDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSGLLVT 81
           205 KTKIEVIKEEERKIREERQEAREEEEQRKQAELALNASSAAAEASS--AQEL----- 254
Db
         82 LIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHPHILFMLLKGYEAPQIALR 141
Qу
           255 LIDTAPVIDAEKTPKV------ATSP-VESPLAPPEVLIM-----GAPK---- 291
Db
        142 CGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADF 201
Qу
                    292 -----TPVATEVDKNADEVEFTK-KDLEVVEDALDTLSKDKNNLVIEKEVIKDI 339
Db
        202 LEQ-----NYDTIFEDYEKL-- 216
Qу
           |:
                                                  ||: :: |
        340 KEEIADYQEDVEELKEAIVAAEKPKDEIKETKGAQRLLKXVNKMITKMDTVVQEIESKES 399
Db
        217 -----LQSENYVTKRQSL---KLLGELILDRHNFAI-MTKYISKPENLKLMMNLL-- 262
Qу
                 Db
        400 EKKAKTLPLEAPRSATETQELDVRKERGEILIDELMDAIKKVKNVPDENRLKLIENILGR 459
        263 -- RDKSPNIQFEAFHVFKVF------VASPHKTQPIVEILLKNQPKLIEFLSSFQKER 312
Qу
              Db
        460 IDTDKDRHIKVE--DVLKVIDIVEKEDGIMSTKQLDELVQLLKKEE--VIELEEKKEKQE 515
        313 TDDEQFADEKNYLIKQIRDLKKTAP 337
QУ
          516 SQQKSFVPPSETLHLESSQQKSTVP 540
RESULT 15
US-09-005-069-70
; Sequence 70, Application US/09005069
; Patent No. 5932470
; GENERAL INFORMATION:
   APPLICANT: FRANK, GLENN R.
   APPLICANT: HUNTER, SHIRLEY WU
   APPLICANT: WALLENFELS, LYNDA
   TITLE OF INVENTION: NOVEL ECTOPARASITE SALIVA PROTEINS
   TITLE OF INVENTION: AND APPARATUS TO COLLECT SUCH PROTEINS
   NUMBER OF SEQUENCES: 107
   CORRESPONDENCE ADDRESS:
     ADDRESSEE: Sheridan Ross P.C.
     STREET: 1700 Lincoln Street, Suite 3500
     CITY: Denver
     STATE: Colorado
     COUNTRY: U.S.A.
    ZIP: 80203
  COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
```

```
;
      COMPUTER: IBM PC compatible
      OPERATING SYSTEM: PC-DOS/MS-DOS
;
      SOFTWARE: PatentIn Release #1.0, Version #1.25
;
    CURRENT APPLICATION DATA:
      APPLICATION NUMBER: US/09/005,069
      FILING DATE:
      CLASSIFICATION:
    PRIOR APPLICATION DATA:
      APPLICATION NUMBER: 08/630,822
      FILING DATE: 11-APR-1996
    ATTORNEY/AGENT INFORMATION:
    NAME: CONNELL, GARY J.
      REGISTRATION NUMBER: 32,020
     REFERENCE/DOCKET NUMBER: 2618-17-C3
    TELECOMMUNICATION INFORMATION:
      TELEPHONE: (303) 863-9700
      TELEFAX: (303) 863-0223
  INFORMATION FOR SEQ ID NO: 70:
;
   SEQUENCE CHARACTERISTICS:
;
     LENGTH: 586 amino acids
;
      TYPE: amino acid
;
      TOPOLOGY: linear
    MOLECULE TYPE: protein
    FEATURE:
      NAME/KEY: Xaa = any amino acid
     LOCATION: 379
US-09-005-069-70
 Query Match 6.2%; Score 105; DB 2; Length 586; Best Local Similarity 20.0%; Pred. No. 0.054;
 Matches 77; Conservative 54; Mismatches 136; Indels 118; Gaps 15;
Qу
         22 KDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSGLLVT 81
            205 KTKIEVIKEEERKIREERQEAREEEEQRKQAELALNASSAAAEASS--AQEL---- 254
Qу
         82 LIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHPHILFMLLKGYEAPQIALR 141
            255 LIDTAPVIDAEKTPKV------ATSP-VESPLAPPEVLIM----GAPK---- 291
Db
        142 CGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADF 201
Qу
                     Db
        292 -----TPVATEVDKNADEVEFTK-KDLEVVEDALDTLSKDKNNLVIEKEVIKDI 339
        202 LEQ-----NYDTIFEDYEKL-- 216
Qу
            |:
                                                   | | | : : : |
Db
        340 KEEIADYQEDVEELKEAIVAAEKPKDEIKETKGAQRLLKXVNKMITKMDTVVQEIESKES 399
        217 -----LQSENYVTKRQSL---KLLGELILDRHNFAI-MTKYISKPENLKLMMNLL-- 262
Qу
                  1:: |: | | | | |:::| || | |: |:|
Db
        400 EKKAKTLPLEAPRSATETQELDVRKERGEILIDELMDAIKKVKNVPDENRLKLIENILGR 459
        263 -- RDKSPNIQFEAFHVFKVF------VASPHKTQPIVEILLKNQPKLIEFLSSFQKER 312
QУ
              460 IDTDKDRHIKVE--DVLKVIDIVEKEDGIMSTKQLDELVQLLKKEE--VIELEEKKEKQE 515
Db
Qу
       313 TDDEQFADEKNYLIKQIRDLKKTAP 337
```

516 SQQKSFVPPSETLHLESSQQKSTVP 540

Search completed: January 7, 2004, 16:45:03 Job time : 29 secs

# GenCore version 5.1.6 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

January 7, 2004, 16:44:17; Search time 21 Seconds Run on:

(without alignments)

1543.278 Million cell updates/sec

US-10-088-872-2 Title:

Perfect score: 1704

1 MKKMPLFSKSHKNPAEIVKI......FADEKNYLIKQIRDLKKTAP 337 Sequence:

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

283308 seqs, 96168682 residues Searched:

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0% Maximum Match 100%

Listing first 45 summaries

PIR 76:\* Database : 1: pir1:\* 2: pir2:\*

3: pir3:\* 4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

,		ક				
Result	_	Query Match	ronath	DΒ	ID	Description
No.	Score	Match :				
1	1376	80.8	341	2	157997	hypothetical calci
2	1063.5	62.4	377	2	T16651	hypothetical prote hypothetical prote
3	1006.5	. 59.1	338	2	T27129	mo25 homolog [impo
4	834.5	49.0	329	2	T50117	hypothetical prote
5	ი85	40.2	305	2	G71441	hypothetical prote
6	632	37.1	348	2 2	B84448 S34681	hypothetical prote
7	4.85	28.5	399 339	2	T33477	hypothetical prote
8	143.5	8.4 7.9	677	2	H64574	DNA topoisomerase
9	134.5 128	7.5	430	2	H64709	hypothetical prote
10: 11	125.5	7.4	298	2	B71685	hypothetical prote
12	125.5	7.4	1642	2	T08880	NMDA receptor-bind hypothetical prote
13	123.5	7.2	1285	2	B72420	nypothetical proce

	400	7.0	1175	2	F64489
14	120	7.0	959	2	T00246
15	118.5	6.7	474	2	S71322
16	1.15	6.7	833	2	T43446
17	113.5	6.6	1411	2	S55123
18	112.5	6.5	725	1	JC5016
19	111.5	6.5	2401	2	T28676
20	111.5		2166	2	G70163
21	111	6.5	2819	2	A90551
22	111	6.5	457	2	C82911
23	109.5	6.4	978	2	A70387
24	109.5	6.4	1830	2	E82909
25	109.5	6.4	695	2	T07283
26	109	6.4		2	S11527
27	109	6.4	1401	2	T18507
28	108.5	6.4	442	2	T50451
29	108.5	6.4	952	2	D64315
30	108.5	6.4	1163		S73254
31	108	6.3	568	2	140055
32	107.5	6.3	483	2.	三90106
33	107.5	6.3	855	2	
34	107.5	6.3	1042	2	G64514
35	107.5	6.3	1726	1	SAZQGM
36	107.5	6.3	1726	2	A45948
3.7	107	6.3	570	2	S68686
28	197	6.3	1173		T43527
39	107	6.3	1727	2	T50073
4.0	106	6.2	474	2	S56748
4.	106	6.2	1295	2	T24587
- 42	105.5	6.2	781	2	T00456
43	105.5	6.2	847	2	A56039
4.4	105.5	6.2	1091	2	
45	105.5	6.2	1619	2	T18499
د ي-					

hypothetical prote DNA polymerase V glutathione symtha hypothetical prote hypothetical prote hyaluronan rec**e**pto rhoptry protein hypothetical prote conserved hypothet hypothetical prote conserved hypothet conserved hypothet hypothetical prote alpha-latrotox**i**n p hypothetical prote hypothetical coile type I restriction replication helica positive trans-act importin beta-1 SU type I restriction major merozoite su major merozoite su phosphoprotein pho sp8 protein - fiss myosin-like coiled glutathione syntha hypothetical prote protein kinase hom GTPase-activating hypothetical prote hypothetical prote

### ALIGNMENTS

```
RESULT 1
157997
hypothetical calcium-binding protein - mouse
C;Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 19-May-2000
C: Accession: I57997
R; Miyamoto, H.; Matsushiro, A.; Nozaki, M.
A; Title: Molecular cloning of a novel mRNA sequence expressed in cleavage stage
A; Reference number: 157997; MUID: 93119656; PMID: 8418809
mouse embryos.
A; Accession: I57997
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: mRNA
A; Residues: 1-341 < RES>
A; Cross-references: GB:S51858; NID:g262933; PIDN:AAB24801.1; PID:g262934
C; Superfamily: Saccharomyces hypothetical protein YKL189w
C; Keywords: calcium binding
```

```
Best Local Similarity 80.7%; Pred. No. 7.9e-85;
                                                                2;
                          32; Mismatches 29; Indels
                                                         Gaps
 Matches 272; Conservative
         4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
           Qу
         1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
Db
         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
           Qv
         61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
Db
        120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
               QУ
        121 CTQQNILFMLLKGYESPETALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
Db
        180 ASDAFATFKOLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
            QУ
        181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
Db
        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
            QУ
         241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQT 300
Db.
         300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTA 336
            301 KLIEFLSKFONDRTEDEOFNDEKTYLVKQTRNLKRAA 337
Tib
                                                        29 37.0
 RESULT: 2
716651
hypothetical protein R02E12.2 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
 C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 18-Feb-2000
 C; Accession: T16651
 R;Leimbach, D.
 submitted to the EMBL Data Library, April 1996
 A; Description: The sequence of C. elegans cosmid R02E12.
 A; Reference number: Z18554
 A; Accession: T16651
 A; Status: preliminary; translated from GB/EMBL/DDBJ
 A; Molecule type: DNA
 A:Residues: 1-377 <LEI>
 A;Cross-references: EMBL:U53337; NID:g1255833; PID:g1255838; PIDN:AAA96187.1;
 GSPDB:GN00028; CESP:R02E12.2
 A; Experimental source: strain Bristol N2; clone R02E12
 C; Genetics:
 A; Gene: CESP: R02E12.2
 A; Map position: X
 A; Introns: 37/3; 146/2; 225/1; 315/3
 C; Superfamily: Saccharomyces hypothetical protein YKL189w
                       62.4%; Score 1063.5; DB 2; Length 377;
   Query Match
   Best Local Similarity 60.5%; Pred. No. 7e-64;
   Matches 211; Conservative 53; Mismatches 68; Indels
                                                       17;
                                                           Gaps
            4 MP-LFSKSHKNPAEIVKILKDNLAILEK------QDKKTDKASEEVSKSLQAM 49
                                                QΥ
```

```
1 MPLLFGKSHKSPADVVKTLREVLTILDKLPPPKLDKDGNIQSDKKYDKALDEVSKNVAMI 60
Db
         50 KEILCGTNEKEPPTE---AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILR 106
                            QΥ
         61 KSFIYGNDSAEPSSEHVVQVAQLAQEVYNANILPMLIKMLPKFEFECKKDVGQIFNNLLR 120
Db
        107 RQIGTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFR 166
            Qу
        121 RQIGTRSPTVEYLGARPEILIQLVQGYSVPDIALTCGLMLRESIRHDHLAKIILYSDVFY 180
Db
         167 DFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKR 226
                    Qу
         181 TFFLYVQSEVFDISSDAFSTFKELTTRHKAIIAEFLDSNYDTFFAQYONLLNSKNYVTRR 240
Db
         227 QSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHK 286
            QУ
         241 ÇSLKLLGELLLDRHNFNTMTKYISNPDNLRLMMELLRDKSRNIQYEAFHVFKVFVANPNK 300
Db
         287 TQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKT 335
             QУ
         301 PKPISDILNRNREKLVEFLSEFHNDRTDDEQFNDEKAYLIKQIQEMKSS 349
 Db
 RESULT 3
 hypothetical protein Y53Cl2A.4 - Caenorhabditis elegans
 C;Species: Caenorhabditis elegans
 C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jul-2000
 C; Accession: T27129
 R; Kershaw, J.; Lennard, N.
 submitted to the EMBL Data Library, September 1997
 A: Reference number: Z20315
 A; Accession: T27129
 A; Status: preliminary; translated from GB/EMBL/DDBJ
 A; Molecule type: DNA
 A; Cross-references: EMBL:Z99277; PIDN:CAB16486.1; GSPDB:GN00020; CESP:Y53C12A.4
 A; Experimental source: clone Y53C12A
  C; Genetics:
  A; Gene: CESP:Y53C12A.4
  A; Map position: 2
  A; Introns: 29/3; 103/3; 136/2; 215/1; 282/3
  C; Superfamily: Saccharomyces hypothetical protein YKL189w
                               Score 1006.5; DB 2; Length 338;
                        59.1%;
    Query Match
                               Pred. No. 3.9e-60;
                        57.2%;
    Best Local Similarity
                                                                    1:
                                                             Gaps
                                                Indels
                            60; Mismatches
                                             78;
    Matches 191; Conservative
            5 PLFSKSHKNPAEIVKILKDNLAILEK-----QDKKTDKASEEVSKSLQAMKEILCGTNEK 59
              Qу
            4 PLFGKADKTPADVVKNLRDALLVIDRHGTNTSERKVEKAIEETAKMLALAKTFIYGSDAN 63
  Db
            60 EPPTEAVAQLAQELYSSGLLVTLIADL:QLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
                 Qy
            64 EPNNEQVTQLAQEVYNANVLPMLIKHLHKFEFECKKDVASVFNNLLRRQIGTRSPTVEYL 123
  Db
           120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
   QУ
```

```
124 AARPEILITLLLGYEOPDIALTCGSMLREAVRHEHLARIVLYSEYFQRFFVFVQSDVFDI 183
Db
        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
           QУ
        184 ATDAFSTFKDLMTKHKNMCAEYLDNNYDRFFGQYSALTNSENYVTRRQSLKLLGELLLDR 243
Db
        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
            QУ
        2-4 HNFSTMNKYITSPENLKTVMELLRDKRRNIQYEAFHVFKIFVANPNKPRPITDILTRNRD 303
Db
        300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK 333
            HEIDER TEHREITER
QУ
        304 KLVEFLTAFHNDRTNDEQFNDEKAYLIKQIQELR 337
Db
RESULT 4
mo25.homolog [imported] - fission yeast (Schizosaccharomyces pombe)
C; Species: Schizosaccharomyces pombe
 C; Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 28-Jul-2000
 R; Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.
 submitted to the EMBL Data Library, February 2000
 A; Reference number: Z25039
 A; Accession: T50117
 A; Status: preliminary; translated from GB/EMBL/DDBJ
 A:Molecule type: DNA
 A; Residues: 1-329 <SEE>
 A:Cross-references: EMBL:AL157734; PIDN:CAB75774.1; GSPDB:GN00066;
 SPDE:SPAC1834.06c
 A; Experimental source: strain 972h(-); cosmid c1834
 C; Genetics:
 A; Gene: SPDB:SPAC1834.06c
 A; Map position: 1
 A; Introns: 34/3; 185/3
 C; Superfamily: Saccharomyces hypothetical protein YKL189w
                       49.0%; Score 834.5; DB 2; Length 329;
   Query Match
                       51.5%; Pred. No. 1.2e-48;
   Best Local Similarity
   Matches 169; Conservative 63; Mismatches 93; Indels
            5 LFSKSHKNPAEIVKILKDNLAILE-KQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTE 64
             QУ
            4 LFNKRPKSTQDVVRCLCDNLPKLEINNDKK--KSFEEVSKCLQNLRVSLCGTAEVFPDAD 61
  Db
           65 AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHPH 124
              QУ
           62 LVSDLSFQIYQSNLPFLLVRYLPKLEFESKKDTGLIFSALLRRHVASRYPTVDYMLAHPQ 121
  Db
          125 ILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAF 184
              QУ
          122 IFPVLVSYYRYQEVAFTAGSILRECSRHEALNEVLLNSRDFWTFFSLIQASSFDMASDAF 181
  Db
           185 ATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFAI 244
              QУ
           182 STFKSILLNHKSQVAEFISYHFDEFFKQYTVLLKSENYVTKPQSLKLLGEILLNRANRSV 241
  Db
```

```
245 MTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIEF 304
             QУ
         242 MTRYISSAENLKLMMILLRDKSKNIQFEAFHVFKLFVANPEKSEEVIEILRRNKSKLISY 301
Db
         305 LSSFQKERTDDEQFADEKNYLIKQIRDL 332
QУ
             11:1 :1:111 11: ::1111
         302 LSAFHTDRKNDEQFNDERAFVIKQIERL 329
Db
RESULT 5
G71441
hypothetical protein - Arabidopsis thaliana
C; Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 03-Aug-1998 #sequence_revision 03-Aug-1998 #text_change 18-Aug-2000
R; Bevan, M.; Bancroft, I.; Bent, E.; Love, K.; Goodman, H.; Dean, C.; Bergkamp,
R.; Dirkse, W.; Van Staveren, M.; Stiekema, W.; Drost, L.; Ridley, P.; Hudson,
S.A.; Patel, K.; Murphy, G.; Piffanelli, P.; Wedler, H.; Wedler, E.; Wambutt,
R.; Weitzenegger, T.; Pohl, T.M.; Terryn, N.; Gielen, J.; Villarroel, R.; De
Clerck, R.; Van Montagu, M.; Lecharny, A.; Auborg, S.; Gy, I.; Kreis, M.; Lao,
N.; Kavanagh, T.; Hempel, S.; Kotter, P.; Entian, K.D.; Rieger, M.; Schaeffer,
 M.; Funk, B.
 Nature 391, 485-488, 1998
 A; Authors: Mueller-Auer, S.; Silvey, M.; James, R.; Montfort, A.; Pons, A.;
 Puigdomenech, P.; Douka, A.; Voukelatou, E.; Milioni, D.; Hatzopoulos, P.;
 Piravandi, E.; Obermaier, B.; Hilbert, H.; Duesterhoft, A.; Moores, T.; Jones,
 U.D.G.; Eneva, T.; Palme, K.; Benes, V.; Rechman, S.; Ansorge, W.; Cooke, R.;
 Derger, C.; Delseny, M.; Voet, M.; Volckaert, G.; Mewes, H.W.; Klosterman, S.;
 Schueller, C.; Chalwatzis, N.
 A; Title: Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of
 Arabidopsis thaliana.
 A; Reference number: A71400; MUID: 98121113; PMID: 9461215
 A; 3tatus: preliminary; nucleic acid sequence not shown; translation not shown
 A; Molecule type: DNA
 A; Cross-references: GB:Z97343; NID:g2245073; PID:e327051; PID:g2245086
 C; Genetics:
 A; Map position: 4COP9-4G3845
 C; Superfamily: Saccharomyces hypothetical protein YKL189w
                                 Score 685; DB 2; Length 305;
                          40.2%;
   Query Match
                                 Pred. No. 1.1e-38;
                          45.9%;
   East Local Similarity
                                                                         2:
                                                    Indels
                                                                 Gaps
   Matches 135; Conservative 68; Mismatches
            41 EVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQI 100
               Qу
             8 ELSKSIRDLKLILYGNSEAEPVAEACAQLTQEFFKADTLRRLLTSLPNLNLEARKDATQV 67
  Dh
           101 FNNILRRQIGTRSPTVEYISAHPHILFMLKGYEAPQIALRCGIMLRECIRHEPLAKITL 160
                              QУ
            68 VANLQRQQVNSRLIAADYLESNIDLMDFLVDGFENTDMALHYGTMFRECIRHQIVAKYVL 127
  Db
           161 FSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDY-EKLLQS 219
                  ÇУ
```

```
128 DSEHVKKFFYYIQLPNFDIAADAAATFKELLTRHKSTVAEFLIKNEDWFFADYNSKLLES 187
Db
        220 ENYVTKRQSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKV 279
             QУ
        188 TNYITRRQAIKLLGDILLDRSNSAVMTKYVSSMDNLRILMNLLRESSKTIQIEAFHVFKL 247
Db
        280 FVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK 333
            Qу
         248 FVANQNKPSDIANILVANRNKLLRLLADIKPDK-EDERFDADKAQVVREIANLK 300
Db
RESULT 6
hypothetical protein At2g03410 [imported] - Arabidopsis thaliana
B84448
C; Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C; Accession: B84448
R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.;
Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell,
C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin,
L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams,
M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver,
G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser,
C.M.; Venter, J.C.
Nature 402, 761-768, 1999
A; Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis
A; Reference number: A84420; MUID: 20083487; PMID: 10617197
 A; Accession: B84448
 A;Status: preliminary
 A, Molecule type: DNA
 A; Residues: 1-348 <STO>
 A; Cross-references: GB:AE002093; NID:g4335758; PIDN:AAD17435.1; GSPDB:GN00139
 C:Genetics:
 A; Gene: At2g03410
 A; Map position: 2
 C; Superfamily: Saccharomyces hypothetical protein YKL189w
                        37.1%; Score 632; DB 2; Length 348;
   Query Match
                        38.7%; Pred. No. 4.4e-35;
   Best Local Similarity
   Matches 133; Conservative 80; Mismatches 117; Indels
            6 LFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASE-----EVSKSLQAMKEILCGTNE 58
                                                  1: :::: :| || |
 QΥ
                  4 LFKNKSRLPGEIVRQTRDLIALAESEEEETDARNSKRLGICAELCRNIRDLKSILYGNGE 63
 Db
           59 KEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEY 118
               QÝ
           64 AEPVPEACLLLTQEFFRADTLRPLIKSIPKLDLEARKDATQIVANLQKQQVEFRLVASEY 123
 D'n
          119 ISAHPHILFMLLKGYEAP-QIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTF 177
              | | | | | | |
 QΥ
          124 LESNLDVIDSLVEGIDHDHELALHYTGMLKECVRHQVVAKYILESKNLEKFFDYVQLPYF 183
 Db
          173 DIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYE-KLLQSENYVTKRQSLKLLGELI 236
              QУ
          184 DVATDASKIFRELLTRHKSTVAEYLAKNYEWFFAEYNTKLLEKGSYFTKRQASKLLGDVL 243
  Db
```

```
237 LDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLK 296
Οv
              244 MDRSNSGVMVKYVSSLDNLRIMMNLLREPTKNIQLEAFHIFKLFVANENKPEDIVAILVA 303
Db
         297 NQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK----KTA 336
Qу
              ]: |:: : : |::| | :| :| |
          304 NRTKILRLFADLKPEK-EDVGFETDKALVMNEIATLSLLDIKTA 346
Db
RESULT 7
S34681
hypothetical protein YKL189w - yeast (Saccharomyces cerevisiae)
C; Species: Saccharomyces cerevisiae
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 19-Apr-2002
C; Accession: S34681; S33963; S38021; S38026
R; Wieman, S.; Voss, H.; Schwagaer, C.; Rupp, T.; Stegemann, J.; Zimmermann, J.;
Grothues, D.; Sensen, C.; Erfle, H.; Hewitt, N.; Banrevi, A.; Ansorge, W.
submitted to the EMBL Data Library, July 1993
A; Description: Sequencing and analysis of 51.5 kilobases on the left arm of
chromosome XI from Saccharomyces cerevisiae reveals 23 open reading frames
including the FAS1 gene.
A; Reference number: S34679
A; Accession: S34681
A; Molecule type: DNA
A; Residues: 1-399 <WIE>
A; Cross-references: EMBL:X74151; NID:g450365; PIDN:CAA52249.1; PID:g395236
A; Experimental source: strain S288C
R; Cheret, G.; Mattheakis, L.C.; Sor, F.
Yeast 9, 661-667, 1993
A; Title: DNA sequence analysis of the YCN2 region of chromosome XI in
Saccharomyces cerevisiae.
A; Reference number: S33960; MUID: 93348778; PMID: 8394042
A, Accession: S33963
A; Molecule type: DNA
A; Residues: 1-399 < CHE>
A; Cross-references: GB:X69765; NID:g296985; PIDN:CAA49422.1; PID:g296989
R; Wiemann, S.; Voss, H.; Schwager, C.; Rupp, T.; Grothues, D.; Sensen, C.;
Stegemann, J.; Zimmermann, J.; Erfle, H.; Hewitt, N.; Ansorge, W.
submitted to the Protein Sequence Database, March 1994
A; Reference number: S37825
A; Accession: S38021
A; Molecule type: DNA
A; Residues: 1-399 <WI2>
A; Cross-references: EMBL: Z28189; NID: g486334; PIDN: CAA82032.1; PID: g486335;
MIPS:YKL189w
A; Experimental source: strain S288C
R; Maia e Silva, A.; Bossier, P.; Vilela, C.; Fernandes, L.; Soares, H.;
Guerreiro, P.; Rodrigues-Pousada, C.
submitted to the Protein Sequence Database, March 1994
A; Reference number: S38024
A; Accession: S38026
A; Molecule type: DNA
A; Residues: 1-399 <MAI>
A; Cross-references: EMBL: Z28189; NID: g486334; PIDN: CAA82032.1; PID: g486335;
MIPS:YKL189w
 A; Experimental source: strain S288C
```

```
C; Genetics:
A; Gene: SGD: HYM1
A; Cross-references: SGD: S0001672
A; Map position: 11L
C; Superfamily: Saccharomyces hypothetical protein YKL189w
                       28.5%; Score 485; DB 2; Length 399;
 Ouery Match
                       33.0%; Pred. No. 3.6e-25;
 Best Local Similarity
 Matches 113; Conservative 75; Mismatches 138; Indels 16; Gaps
                                                                      6;
           7 FSKSHKNPAEIVKILKDNLAILEK----QDKKTDKASEEVSKSLQAMKEILCGTNEKEPP 62
Qу
                                      : |: | |:: ::: : |
          16 WKKNPKTPSDYARLIIEQLNKFSSPSLTQDNKR-KVQEECTKYLIGTKHFIVGDTDPHPT 74
Db
          63 TEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 122
QУ
             : ||:|:::
          75 PEAIDELYTAMHRADVFYELLLHFVDLEFEARRECMLIFSICLGYSKDNKFVTVDYLVSQ 134
Db
         123 PHILFMLLKGYE-----APQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELS 175
Qу
                                | :::|: |
         135 PKTISLMLRTAEVALQQKGCQDIFLTVGNMIIECIKYEQLCRIILKDPQLWKFFEFAKLG 194
Db
         176 TFDIASDAFATFKDLLTRHKVLVA-DFL--EQNYDTIFEDYEKLLQSENYVTKRQSLKLL 232
Qу
                           1: ::::
         195 NFEISTESLQILSAAFTAHPKLVSKEFFSNEINIIRFIKCINKLMAHGSYVTKRQSTKLL 254
7h
         233 GELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVE 292
               255 ASLIVIRSNNALMNIYINSPENLKLIMTLMTDKSKNLQLEAFNVFKVMVANPRKSKPVFD 314
Db
         293 ILLKNOPKLIEFLSSFOKERTDDEQFADEKNYLIKQIRDLKK 334
QУ
             ]]:]]: |]: : : | : : | | ||: ::::: | | :
         315 ILVKNRDKLLTYFKTFGLD-SQDSTFLDEREFIVQEIDSLPR 355
Dò
RESULT &
T33477
hypothetical protein T27C10.3 - Caenorhabditis elegans
C; Species: Caenorhabditis elegans
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C; Accession: T33477
R; Zhu, H.J.; Graves, T.; Hawkins, M.
submitted to the EMBL Data Library, October 1998
A; Description: The sequence of C. elegans cosmid T27C10.
A; Reference number: Z21354
A: Accession: T33477
A; Status: preliminary; translated from GB/EMBL/DDBJ
A: Molecule type: DNA
A; Residues: 1-339 <ZHU>
A; Cross-references: EMBL:AF098504; PIDN:AAC67411.1; GSPDB:GN00019; CESP:T27C10.3
A; Experimental source: strain Bristol N2; clone T27C10
C:Genetics:
A; Gene: CESP: T27C10.3
A; Map position: 1
A; Introns: 72/3; 120/3; 233/3; 295/1
                         8.4%; Score 143.5; DB 2; Length 339;
  Query Match
```

```
Best Local Similarity 19.3%; Pred. No. 0.02;
 Matches 38; Conservative 50; Mismatches 76; Indels
                                                            33; Gaps 4;
         159 ILFSNOFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQ 218
Qу
                                   | | : : :
                              | | :
                                                   : | : : |
             :: :|:||
                         ----FDVIQGTFDTLQIIFFTNHESANNFIKNNLPRFMQTLHKLIA 150
         100 LMNTNKFRD---
Db
         219 SENYVTKROSLKLIGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFK 278
Qy
               151 CSNFFIQAKSFKFLNELFTAQTNYETRSLWMAEPAFIKLVVLAIQSNKHAVRSRAVSILE 210
Db
         279 VFVASPHKTOPIVEILLKNOPKLIEFL------SSFQKERTDDEQFAD----- 320
Qу
                                                 1 : 11
             :|::|::|:||
                                                       | |:|
         211 IFIRNPRNSPEVHEFIGRNRNVLIAFFFNSAPIHYYQGSPNEKE---DAQYARMAYKLLN 267
         321 ---EKNYLIKQIRDLKK 334.
Qу
               :::: : : ::: ::
Db
         268 WDMQRPFTQEQLQDFEE 284
RESULT 9
H64574
DNA topoisomerase I - Helicobacter pylori (strain 26695)
C; Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text change 08-Oct-1999
C; Accession: H64574
R; Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.;
Pleischmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.;
Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, K.; Fitzegerald,
L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback,
T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.;
Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A; Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.;
Fraser, C.M.; Venter, J.C.
A; Title: The complete genome sequence of the gastric pathogen Helicobacter
pylori.
A; Reference number: A64520; MUID: 97394467; PMID: 9252185
A; Accession: H64574
A; Status: preliminary; nucleic acid sequence not shown; translation not shown
A; Molecule type: DNA
A, Residues: 1-677 < TOM>
A; Cross-references: GB: AE000559; GB: AE000511; NID: g2313536; PIDN: AAD07502.1;
PID:q2313542; TIGR:HP0440
C; Superfamily: DNA topoisomerase I
                         7.9%; Score 134.5; DB 2; Length 677;
  Query Match
                        21.6%; Pred. No. 0.19;
  Best Local Similarity
  Matches 88; Conservative 58; Mismatches 134; Indels 127; Gaps
           7 FSKSHKNPA-EIVKILKDNL------AILEKQDKK---TDKASEEVSKSLQAMKE 51
QУ
                  11 | : : | | | | | |
                                          :: |: ||
                                                      222 FKFKDKNEASQFLKDLKDGLGSMSVLVSLKESLSNKEPKKPFTTSKLLSQASKSLKI--- 278
Db
          52 ILCGTNEKEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGT 111
QУ
```

```
279 ------PTKE:IAQLAQKLFEAGLITYHRTDSEFLSPEYLKEHEVFFEPIY----- 322
Db
        112 RSPTV----EYIS------AHPHILFMLLKGYEAPQIALRCGIMLRECIRHE 153
Qу
                                    |:| || :
        323 --PSVYQYREYKAGKNSQAEAHEAIRITHPHALKDLEKVCSDAKISEELALKLYQLIYTN 380
Db
        154 PL---AKIILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIF 210
QУ
            361 TICSQSRNALY-NQYDCIFK-----IKSESFKLSFKLLKEKGFLEIEELIQGKEEIN 431
ΔĎ
         211 EDYEKLLQSENYVTKRQSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQ 270
Qу
             432 RE-EQESEIENFSLKENDSVPLKEVFIKK----IEKPSPKPYKESAFIPLLESEG---- 481
Dò
         271 FEAFHVFKVFVASPHKTQPIVEILLKNQ------PKLIEFLSSFQKERTDD- 315 :
QУ
                                         : :| :| :| ::
                    : | '' :: | | :
                 ----IGRPSTYASFLDLLLKRKYISIDTKTNAITPTSQGLEVISFFKKDKEVDF 531
Db
         316 -----EQF-----ADEKNYLIKQIRDLKKTA 336
QУ
                                | : :::: | || ||
                          : ] ]
         532 IALTSKDKSKLGNTTKQFEECLDLIMRGEASYEKFMLEVISKLKSTA 578
Db
RESULT 10
H64709
hypothetical protein HP1520 - Helicobacter pylori (strain 26695)
C; Species: Helicobacter pylori
C;Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999
C; Accession: H64709
R; Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.;
Fleaschmann, R.D.; Ketchum, K.A.; Klenk, H.P.; Gill, S.; Dougherty, B.A.;
Nelson, K.; Quackenbush, J.; Zhou, L.; Kirkness, E.F.; Peterson, S.; Loftus, B.;
Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenney, K.; Fitzegerald,
L.M.; Lee, N.; Adams, M.D.; Hickey, E.K.; Berg, D.E.; Gocayne, J.D.; Utterback,
T.R.; Peterson, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.;
Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A; Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.;
Fraser, C.M.; Venter, J.C.
A: Title: The complete genome sequence of the gastric pathogen Helicobacter
A; Reference number: A64520; MUID: 97394467; PMID: 9252185
A, Accession: H64709
A; Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A; Residues: 1-430 < TOM>
A:Cross-references: GB:AE000650; GB:AE000511; NID:g2314700; PIDN:AAD08565.1;
PID:g2314705; TIGR:HP1520
C; Superfamily: Helicobacter pylori hypothetical protein HP1520
                        7.5%; Score 128; DB 2; Length 430;
  Oliery Match
  Best Local Similarity 20.9%; Pred. No. 0.29;
  Matches 82; Conservative 73; Mismatches 135; Indels 102; Gaps
           7 FSKSHKNPAEI----VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPP 62
QУ
             60 FYPNRKSKIEIEFNGEKILKENVAVFHSYDE--EFSSEDSVTTFMAKSDL-----KQQY 111
Db
```

```
63 TEAVAQLAQELYSSGLLVTL--IA------DLQLIDFEGKKDVTQIFNNILR----- 106
QУ
               : :| :| | | | | | | | | |
                                       112 DNILLELEKE--KKALLKSLRDIASGFDYEEEIKTIKNEKNKSFYEILDNHLTEIESSEK 169
Db
         107 -----RQIGTRSPTV-EYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKII 159
Qу
                 170 HYSFKYRDIFDGSKKVKDFVNKHHDLIEQYFNKYQ---
                                                         ---ELLSQSK 211
Db
         160 LF-----SNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQ----- 204
QУ
                     212 IFKHMNSGDFGTNHADDLKKALENNRFFKANHSLKIAGEEITNYQKL-SDIFENEKNRIL 270
Db
         205 NYDTIFEDYEKULQSENYVTKRQSLKLLGELI-----LDRHNF--AIMTKYISKP 252
QΥ
            | : | ::|: | : : | : : |: :
         271 NNEELKESFDKI---EKVINANKELKAFKDAISKDNTLLTEFLDYDSFRKKVLFSYLKQV 327
Db
         253 -ENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKE 311
Qу
             328 IQNVKSLVNLYREKKPEIE----EIIKQASKDQKEWESVIEIF--NQRFLVPFKVELQNQ 381
Db
         312 R----TDDEQ----FADEKNYLIKQIRDLKK 334
ŰУ
            382 KDILLNKDAAQFRFIFSDDNQDMNVQKEDLQK 413
Db
RESULT 11
571585
hypothetical protein RP295 - Rickettsia prowazekii
C;Species: Rickettsia prowazekii 🦤
C/Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
C; Accession: B71685
R; Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Sicheritz-Ponten, T.;
Alsmark, U.C.M.; Podowski, R.M.; Naeslund, A.K.; Eriksson, A.S.; Winkler, H.H.;
Kurland, C.G.
Nature 396, 133-140, 1998
A; Title: The genome sequence of Rickettsia prowazekii and the origin of
mitochondria.
A; Reference number: A71630; MUID: 99039499; PMID: 9823893
A; Accession: B71685
A; Status: preliminary; nucleic acid sequence not shown; translation not shown
A; Molecule type: DNA
A:Residues: 1-298 <AND>
A; Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAA14756.1;
PID:g3860856; GSPDB:GN00081
A, Experimental source: strain Madrid E
C: Genetics:
A:Gene: RP295
                        7.4%; Score 125.5; DB 2; Length 298;
  Query Match
  Best Local Similarity 20.1%; Pred. No. 0.27;
  Matches 62; Conservative 57; Mismatches 114; Indels
                                                         75; Gaps
          73 LYSSGLLVTLIADLQLIDFEGKKDVTQ------IFNNILRRQIGTRS 113
QУ
                                                      :: |: |
            |: |:||: ::|:: :||
           6 LFIQLLIVTSLVKAEIIEVDSLNKITQDFKVNYNKNYLPQDLLVVTVLDKFLFKSFGV-- 63
Db
```

```
114 PTVEYISAHPHILFMLLKGY--EAPQIALRCGIMLRECIRHEPLAKIILFSNQFR---- 166
Qу
            64 PIGEYIDQHRYLALAPLFSHINKNPKIIY------ITQLILTNNSYKKELQE 109
Db
        167 -DFFKYV-ELSTFDI----ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSE 220
Qy
             110 SDFPNFVNEMSNSQIPIIAVNNGFTGNFNNIPKFEIWFADYLKKNF---YIDFSKSFPNN 166
Db
         221 NYVTKRQSLKLLGELILDRHNFAIMTKYISKPENL---KLMMNLLRDKSPNIQFEAFHVF 277
Qу
                   167 NYI-----IFNNLDSFDNTYPVFYKGILTSNNIPASKVILNFL-----IQINFIPKC 213
Db
         278 KVFVASPHKTQPIVEILLKNQPKLIEFLSSF--QKERTDDEQFADEKNY-----LIKQI 329
QУ
                                          | :||: :| |
             214 FILISSSRELLRSMEFQLNNYSSNILFIGYHYNNKSISDDKDYKDIAYYTKMINDLIPQI 273
Db
         330 RDLKKTAP 337
QУ
              274 NKLKRNNP 281
Db
RESULT 12
T08880
NMDA receptor binding protein yotiao - human
C; Species: Homo sapiens (man)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C/Accession: T08880
R; Lin, J.W.; Wyszynski, M.; Madhavan, R.; Sealock, R.; Kim, J.U.; Sheng, M.
J. Neurosci. 18, 2017-2027, 1998
A; Title: Yotiao, a novel protein of neuromuscular junction and brain that
interacts with specific splice variants of NMDA receptor subunit NR1.
A; Reference number: Z16511; MUID: 98151389; PMID: 9482789
A; Accession: T08880
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: mRNA
A; Residues: 1-1642 <LIN>
A; Cross-references: EMBL: AF026245; NID: g2623067; PIDN: AAB86384.1; PID: g2623068
C:Genetics:
A; Map position: 7q21-22
C; Keywords: brain; cerebral cortex; coiled coil; neuromuscular junction;
skeletal muscle
                        7.4%; Score 125.5; DB 2; Length 1642;
  Query Match
  Best Local Similarity 20.2%; Pred. No. 2.4;
  Matches 77; Conservative 73; Mismatches 117; Indels 115; Gaps
          18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77
QУ
             ||:
         664 IEKLKDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE---- 710
Db
          78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI------GTRSPTVEYISAHPHI 125
QУ
             Carp III profess II I hereby and II also the
         711 --ISKLKDLQQSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 766
Db
         126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
Õλ
                                              | :: | :|
                :1 |
                                        -----NDLQEKFAQLEAEN-SILKDEKK 797
          767 LEKOMKEKE----
Db
```

```
186 TFKDLLTRH------KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237
Qу
            798 TLEDMLKIHTPVSQEERLIFLDSIKSKSKDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 857
Db
                                                   YISKPENLKLMMNLLRD 264
         238 DRHNFAIMTK----
QУ
                                                   | | | : | :: | |
             1: 1:
         858 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNKQELEYKSKLKALNEELHLQRI 917
Db
         265 KSPNIQFEA--FHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320
QУ
               918 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVXEKDTTELMEKLEVTKREKLELSQRLSDL 974
Dh
         321 -----EKNYLIKQIRDLK 333
QУ
                    | ::| :::: ||
         975 SEOLKPKPGEISFLNEEVKSLK 996
Db
RESULT 13
B72420
hypothetical protein TM0088 - Thermotoga maritima (strain MSB8)
C; Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C; Accession: B72420
P; Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft,
D.H.; Hickey, E.K.; Peterson, J.D.; Nelson, W.C.; Ketchum, K.A.; McDonald, L.;
Utterback, T.R.; Malek, J.A.; Linher, K.D.; Garrett, M.M.; Stewart, A.M.;
Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, D.; Heidelberg, J.;
Sutton, G.G.; Fleischmann, R.D.; White, O.; Salzberg, S.L.; Smith, H.O.; Venter,
J.C.; Fraser, C.M.
Nature 399, 323-329, 1999
A; Title: Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima.
A; Reference number: A72200; MUID: 99287316; PMID: 10360571
A:Accession: B72420
A; Status: preliminary
A; Molecule type: DNA
A:Residues: 1-1285 <ARN>
A:Cross-references: GB:AE001695; GB:AE000512; NID:g4980569; PIDN:AAD35182.1;
PID:g4980577; TIGR:TM0088
A; Experimental source: strain MSB8
C; Genetics:
A; Gene: TM0088
                         7.2%; Score 123.5; DB 2; Length 1285;
  Query Match
  Best Local Similarity 21.5%; Pred. No. 2.4;
Matches 86; Conservative 78; Mismatches 129; Indels 107; Gaps
           1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQD-----KKT----DKASEEV----SKS 45
QУ
             556 LKVAMLSGKEEEN----VQKAAEELQIISSEERIIRFVKKTENVPIDKAKNVVLQLYSVS 611
Db
          46 LQAMKEILCGTNEKEPPTEAVAQLAQELYSSGL-----LVTLIAD-- 85
QУ
                                                           : :: :
             612 IEELGNELVVIGERE-EVEKAADLLQKIFSSEVEISRDFVKLPSWIDEQEKLLEVVKNSA 670
Db
          36 --- LQI.ID---- FEGKKD---- VTQI FNNILRRQIG-- TRSPTVEYI--- SAHPHILFML 129
QУ
```

```
| | | | | :
         671 GITYEILDGVVYFEGTKENVEKAKELFSDIVEK-LGEVRKEETVEFLEVNSSFPVDEFIN 729
ďŒ
         130 LKGYEAPQIALRCGIMLRECIRHEPLAKIIL-----FSNQFRDFF----KYVELST 176
Оy
                                                1: :| ||
                            -----CFSLDQLGLLVLKGSSEAVEDLSSMYRSFFERHQKIVKENV 780
         730 LSGKLYPDVT--
Db
         177 FD---IASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLG 233
Qy
            781 FDRLMLEVPSGFSFEEFKTFLEVLVPEVKQ----VVYLDKLNLLLVEVPVSQSERVKSLL 836
Db
         234 ELILDRHNFAIMTKYIS------KPENL-KLMMNLLRDKSPNIQFEAF-HVFKVFVAS 283
Qу
            : : :
                                   837 DTFLKKEEAVSEKKAVKSVTIPSGVNPDELSSYLKKLLR----NVEITVFPNMGQMIVEG 892
dQ
         284 P-HKTQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEK 322
QУ
            893 PENEVEKAVELVEAEKEKIV-----LKERKDYVKVSDGK 926
RESULT 14
F64489
hypothetical protein MJ1519 - Methanococcus jannaschii
C; Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text change 21-Jul-2000
C:Addession: F64489
R; Bult, C.J.: White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.;
Blake, J.A.; FitzGerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.;
Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness,
N.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghagen,
N.S.M., Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley,
U.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.;
Hurst, M.A.
Science 273, 1058-1073, 1996
A: Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.;
Woese, C.R.; Venter, J.C.
A, Title: Complete genome sequence of the methanogenic archaeon, Methanococcus.
jannaschii.
A; Reference number: A64300; MUID: 96337999; PMID: 8688087
A; Accession: F64489
A; Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A; Residues: 1-1175 <BUL>
A; Cross-references: GB: U67593; GB: L77117; NID: g2826427; PIDN: AAB99538.1;
PID:q1500409; TIGR:MJ1519
C:Genetics:
A; Map position: FOR1494096-1497623
                        7.0%; Score 120; DB 2; Length 1175;
  Query Match
  Best Local Similarity 21.5%; Pred. No. 3.6;
 Matches 76; Conservative 58; Mismatches 131; Indels
           7 FSKSHKNPAEIVKILKD-NLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEA 65
QУ
             232 FNKFREENQDFDKYLTDENIAFRPHVMKKFDEFAENIKKVIAELE----GSKYKYPGLPG 287.
Db
          66 VAOLAOELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHPHI 125
Qу
```

```
.:| ||:| :::| :: : | :|:
                   288 V-----LYFLGMEDAYSRYIELWKNEGEKGEEKLYNALI-ESLENRKENLEF----- 333
Db
         126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFK----YVELSTFDIA- 180
Oy
                                  ----GITKKVIDKFIAQKEEFREFLKNYAVYYELSAFKLEK 370
Db
         181 ----SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSL----- 229
Qу
                                           : :|::: |
                   371 IKEQYEKEFINLDNIIKNPYILVED-LKEN-----DSFERIIFEELDSWERRRLGDKFNP 424
Db
         230 -----KLLGELILDRH----NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAF---- 274
QУ
                    425 YSPYRVRALLVE-ILKRHLSSGNTTISTK-----DLKDFFEKMDKDIVKITFDEFLRII 477
Db
         275 HVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIK 327
Qу
              478 EEYKDIIS--EKVEIVKKEVKNNENKEIIELFTLKEIREYEEIIENTINYLLK 528
Db
RESULT 15
900246
DNA polymerase V - fission yeast (Schizosaccharomyces pombe)
C; Species: Schizosaccharomyces pombe
C;Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 31-Jan-2000
C; Accession: T00246; T39442
R;Shimizu, K.
Submitted to the EMBL Data Library, March 1998
RyDescription: S.pombe homolog of S.cerevisiae DNA polymerase V.
AgReference number: Z14129
A:Accession: T00246
A)Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: mRNA
A; Residues: 1-959 <SHI>
A; Cross-references: EMBL: AB012696; NID: d1224325; PIDN: BAA32046.1; PID: d1033008
R; Hiang, Z.; Aves, S.; Lyne, M.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, March 1998
A; Reference number: Z21854
A; Accession: T39442
A; Status: preliminary; translated from GB/EMBL/DDBJ
A: Molecule type: DNA
A, Residues: 1-959 <LYN>
A; Cross-references: EMBL: AL022305; PIDN: CAA18436.1; GSPDB: GN00067;
SPDB:SPBC14C8.14c
A; Experimental source: strain 972h-; cosmid c14C8
C:Genetics:
A; Cene: pol5+; SPBC14C8.14c
A; Map position: 2
A; Introns: 66/3
                         7.0%; Score 118.5; DB 2; Length 959;
  Query Match
  Best Local Similarity 20.5%; Pred. No. 3.5;
  Matches 80; Conservative 63; Mismatches 135; Indels 113; Gaps
           9 KSHKN-----PAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPP 52
\Omega Y
                  522 KSPKNNLLISMDESVIEIVQKSLSVLHKVTKKIDKKAQHL-QQLNAF----- 567
Db
```

Qy		TEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI-	
Db	568	QLLYSLVLLQVYAGDTDSIDVLEDIDNCYSKVFNKKSKRESTSNEPTAMEIL	619
Qу		SAHPHILFMLLKGYEAPQIALRCGIMLRECI	
Db	620	TEVMLSLLSRPSLLLRKLVDMLFTSFSEDMNRESIHLICDVLKAKESVKDSEGMFAGEEV	679
Qу		RHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIF : : :    : :: :    : ::: :    :	
рр	680	EEDAFGETEMDEDDFEEIDTDEIEEQSDWEMISNQDASDNEELERKLDKVL	730
QУ		EDYEKLLQSENYVTKRQSLKLLGELILDRHNFAIMTKYISKPENLKLMMNIL	
ФŒ	73 <u>1</u>	EDADAKVKDEESSEEELMNDEQMLALDEKLAEVFRERKKASNKEKKKNAQ	780
QŸ		RDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIE	
Do	781	ETKQQIVQFKVKVIDLIDNYYKTQPNNGLGFEFLIPLLEMILKTKHKVLEEKGQAV	836
QУ	304	FLSSFQKERTDDEQFADEKNYLIKQIRDL 332	
dG	837	FRNRLSKLKWTEEK-PSSKNVLEALKKVHVL 866	٠

Search completed: January 7, 2004, 16:46:10 Tob time: 32 secs

## GenCore version 5.1.6 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

January 7, 2004, 16:46:15; Search time 35 Seconds Run on:

(without alignments)

1940.251 Million cell updates/sec

US-10-088-872-2 Title:

Perfect score: 1704

1 MKKMPLFSKSHKNPAEIVKI......FADEKNYLIKQIRDLKKTAP 337 Sequence:

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

747907 seqs, 201509753 residues Searched:

Total number of hits satisfying chosen parameters: 747907

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Published\_Applications\_AA:\* Database :

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*

3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*

4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*

5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*

/cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*

/cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*

/cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*

/cgn2\_6/ptodata/2/pubpaa/US09A\_PUBCOMB.pep:\* 9:

10: /cgn2 6/ptodata/2/pubpaa/US09B PUBCOMB.pep:\*

11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*

12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\* 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep:\*

14: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*

15: /cgn2 6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*

/cgn2\_6/ptodata/2/pubpaa/US10\_NEW PUB.pep:\* 16: /cgn2\_6/ptodata/2/pubpaa/US60 NEW PUB.pep:\* 17:

/cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

용 Result Query Score Match Length DB No.

Description

Sequence 1, Appli US-10-025-730-1 337 100.0 1704 1 Sequence 5, Appli US-10-239-079-5 496 12 81.0 2 1381 Sequence 6, Appli US-10-239-079-6 552 12 81.0 1381 3 Sequence 3, Appli US-10-025-730-3 341 15 80.8 1376 4 Sequence 4, Appli US-10-025-730-4 339 15 1109 65.1 Sequence 5, Appli US-10-025-730-5 377 15 62.4 6 1063.5 Sequence 32324, A US-10-029-386-32324 820 12 7.5 7 128.5 Sequence 11, Appl US-10-080-608A-11 3878 12 7.5 128.5 8 Sequence 4, Appli US-10-171-311-4 3899 15 7.5 9 128.5 Sequence 2, Appli US-10-171-311-2 3907 15 7.5 10 128.5 Sequence 100, App US-10-370-685-100 12 3911 128.5 7.5 11 Sequence 8, Appli US-10-171-311-8 15 7.5 3917 128.5 12 Sequence 6, Appli US-10-171-311-6 3925 15 7.5 128.5 13 Sequence 47959, A US-09-864-761-47959 660 9 7.3 14 125 Sequence 18, Appl 709 12 US-10-023-634-18 6.8 116.5 15 Sequence 77, Appl US-10-023-634-77 709 12 6.8 116.5 16 Sequence 80, Appl US-10-023-634-80 12 724 116.5 6.8 17 Sequence 6, Appli US-10-304-095-6 2184 12 6.6 113 18 Sequence 79, Appl US-10-023-634-79 725 6.5 111.5 19 Sequence 7117, Ap US-10-032-585-7117 769 12 108 6.3 20 Sequence 15, Appl US-10-256-250-15 12 709 6.2 106.5 21 Sequence 14, Appl US-10-023-634-14 709 12 6.2 106.5 22 US-09-978-309A-47 Sequence 47, Appl 725 11 6.2 23 106.5 Sequence 14, Appl US-10-256-250-14 725 12 6.2 24 106.5 Sequence 78, Appl US-10-023-634-78 725 12 6.2 106.5 25 Sequence 16, Appl: US-10-023-634-16 6.2 646 1.2 106 26 Sequence 6440, Ap US-10-369-493-6440 6.2 1295 12 106 27 Sequence 6328, Ap US-10-369-493-6328 6.2 1091 28 205.5 Sequence 30, Appl US-10-071-751-30 14 105 6.2 586 29 Sequence 25, Appl . US-10-369-493-25 30 104 6.1 859 12 Sequence 4, Appli US-09-350-874-4 2.03,5 6.1 245 9 31 Sequence 4, Appli: US-10-106-989-4 245 15 6.1 32 103.5 Sequence 7776, Ap US-10-032-585-7776 6.0 1038 12 103 33 Sequence 168, App US-09-745-763-168 9 677 102.5 6.0 34 Sequence 19, Appl US-10-080-608A-19 6.0 975 12 102.5 35 Sequence 108, App US-10-370-685-108 6.0 975 12 102.5 36 Sequence 7365, Ap US-10-032-585-7365 12 6.0 1106 37 102.5 Sequence 446, App. US-10-295-027-446 1416 12 38 102.5 6.0 Sequence 73, Appl. US-09-820-843A-73 1805 11 6.0 39 102 Sequence 22734, A US-10-369-493-22734 1837 12 102 6.0 40 Sequence 2070, Ap US-10-369-493-2070 1957 12 5.0 41 102 Sequence 82, Appl US-10-146-473-82 1388 15 6.0 42 101.5 Sequence 7646, Ap 1881 US-10-032-585-7646 12 6.0 101.5 43 Sequence 4, Appli 9 US-09-402-100-4 1338 5.9 201 44 Sequence 44, Appl 12 US-10-205-219-44 1230 5.9 100.5 45

### ALIGNMENTS

### RESULT 1

US-10-025-730-1

<sup>;</sup> Sequence 1, Application US/10025730

<sup>;</sup> Publication No. US20030045466A1

<sup>;</sup> GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

```
Guegler, Karl J.
  APPLICANT:
            Corley, Neil C.
  APPLICANT:
            Gorgone, Gina A.
  APPLICANT:
  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
  FILE REFERENCE: PF-0635 US
  CURRENT APPLICATION NUMBER: US/10/025,730
  CURRENT FILING DATE: 2001-12-18
  PRIOR APPLICATION NUMBER: US/09/190,965
  PRIOR FILING DATE: 1998-11-13
  NUMBER OF SEQ ID NOS: 5
  SOFTWARE: PERL Program
 SEQ ID NO 1
   LENGTH: 337
   TYPE: PRT
   ORGANISM: Homo sapiens
   FEATURE: -
   OTHER INFORMATION: 3734805
US-10-025-730-1
                                              Length 337;
                             Score 1704; DB 15;
                      100.0%;
 Query Match
                             Pred. No. 3.1e-147;
                      100.0%;
 Best Local Similarity
                           0; Mismatches
                                           0; Indels
                                                          Gaps
 Matches 337; Conservative
          1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKE 60
QΥ
           MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKE 60
DЪ
         51 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
QУ
            61 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
Db
        121 AHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
Qу
            121 AHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
Db
        181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
Çу
            SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
Db
        241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPK 300
Qу
            241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPK 300
Db
        301 LIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
QV
            301 LIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
Db
RESULT 2
US-10-239-079-5
; Sequence 5, Application US/10239079
  Publication No. US20030148446A1
  GENERAL INFORMATION:
   APPLICANT: Merck Patent GmbH
   TITLE OF INVENTION: ANIC-BP1-ligand
   FILE REFERENCE: ANIC-BP-1-ligand
   CURRENT APPLICATION NUMBER: US/10/239,079
```

```
CURRENT FILING DATE: 2002-09-19
  NUMBER OF SEQ ID NOS: 8
  SOFTWARE: PatentIn Ver. 2.1
 SEO ID NO 5
   LENGTH: 496
   TYPE: PRT
   ORGANISM: Artificial Sequence
   FEATURE:
   OTHER INFORMATION: Description of Artificial Sequence: Gal4-ANIC-BP-1
   OTHER INFORMATION: fusion protein
US-10-239-079-5
                     81.0%; Score 1381;
                                       DB 12;
                                              Length 496;
 Query Match
                     81.0%; Pred. No. 1.6e-117;
 Best Local Similarity
                                                          Gaps
                                              Indels
 Matches 273; Conservative
                          31; Mismatches
          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
QУ
           156 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 215
Db
         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
QУ
           216 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 275
Db
        120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
QУ
               276 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 335
Db
        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
QУ
            336 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 395
Do
        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
QУ
            396 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTOPILDILLKNQA 455
Db
        300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTA 336
Qу
            456 KLIEFLSKFQNDRTEDEQFNDEKTYLVKQIRDLKRPA 492
Db
RESULT 3
US-10-239-079-6
; Sequence 6, Application US/10239079
 Publication No. US20030148446A1
  GENERAL INFORMATION:
  APPLICANT: Merck Patent GmbH
  TITLE OF INVENTION: ANIC-BP1-ligand
  FILE REFERENCE: ANIC-BP-1-ligand
  CURRENT APPLICATION NUMBER: US/10/239,079
  CURRENT FILING DATE: 2002-09-19
  NUMBER OF SEQ ID NOS: 8
  SOFTWARE: PatentIn Ver. 2.1
  SEQ ID NO 6
   LENGTH: 552
   TYPE: PRT
   ORGANISM: Artificial Sequence
```

```
FEATURE:
ï
   OTHER INFORMATION: Description of Artificial Sequence: LexA-ANIC-BP-1
   OTHER INFORMATION: fusion protein
US-10-239-079-6
                      81.0%; Score 1381; DB 12;
                                              Length 552;
 Query Match
                      81.0%; Pred. No. 1.9e-117;
 Best Local Similarity
                                                                 2.;
 Matches 273; Conservative
                           31; Mismatches
                                              Indels
                                                          Gaps
          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
Qy
            212 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 271
Db
         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
QУ
            272 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 331
Db
        120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
Qу
               332 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 391
Db
        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
Qу
            392 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 451
Db
        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299:
QУ
            452 HNFTIMTKYISKPENIKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 511
Db
        300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTA 336
\Omega Y
            512 KLIEFLSKFQNDRTEDEQFNDEKTYLVKQIRDLKRPA 548
Db
RESULT 4
US-10-025-730-3
; Sequence 3, Application US/10025730
 Publication No. US20030045466A1
 CENERAL INFORMATION:
  APPLICANT: Tang, Y. Tom
  APPLICANT: Guegler, Karl J.
  APPLICANT: Corley, Neil C.
  APPLICANT: Gorgone, Gina A.
  TITLE OF INVENTION: CALCIUM BINDING PROTEIN
  FILE REFERENCE: PF-0635 US
  CURRENT APPLICATION NUMBER: US/10/025,730
  CURRENT FILING DATE: 2001-12-18
  PRIOR APPLICATION NUMBER: US/09/190,965
  PRIOR FILING DATE: 1998-11-13
  NUMBER OF SEQ ID NOS: 5
  SOFTWARE: PERL Program
 SEC ID NO 3
   LENGTH: 341
   TYPE: PRT
   ORGANISM: Mus sp.
   FEATURE: -
```

OTHER INFORMATION: q262934

```
Score 1376; DB 15;
                                                                                                         Length 341;
                                                 80.8%;
   Query Match
                                                 80.7%; Pred. No. 2.8e-117;
   Best Local Similarity
                                                                                                                                    Gaps
                                                                                                                                                    2:
                                                                                                                             4;
   Matches 272; Conservative 32; Mismatches
                                                                                                         Indels
                                                                                                29;
                      4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
                           Qу
                      1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
Db
                     60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
                           Qу
                     61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQ1FNNILRRQIGTRTPTVEYI 120

d\mathbf{G}

                   1.20 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
                                   QУ
                   121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
Db
                   180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILLR 239
                           MINITER NAMED AND ASSESSED ASSESSEDA ASSESSED ASSESSED ASSESSED ASSESSED ASSESSED ASSESSED ASSESSEDA ASSESSED ASSESSED ASSESSED ASSESSED ASSESSED ASSESSED ASSESSEDA ASSESSED ASSESSED ASSESSED ASSESSED ASSESSED ASSESSED ASSESSEDA
ÇΥ
                    181 ASDAFATFKOLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
Db
                    240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLENQP 299
QУ
                            241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQT 300
Db
                    300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTA 336
 \bigcirc \nabla
                            301 KLIEFLSKFONDRTEDEQFNDEKTYLVKQIRNLKRAA 337
 RESULT 5
 US-10-025-730-4
 ; Sequence 4, Application US/10025730
     Publication No. US20030045466A1
     GENERAL INFORMATION:
       APPLICANT: Tang, Y. Tom
       APPLICANT: Guegler, Karl J.
       APPLICANT: Corley, Neil C.
       APPLICANT: Gorgone, Gina A.
       TITLE OF INVENTION: CALCIUM BINDING PROTEIN
        FILE REFERENCE: PF-0635 US
       CURRENT APPLICATION NUMBER: US/10/025,730
        CURRENT FILING DATE: 2001-12-18
        ERIOR APPLICATION NUMBER: US/09/190,965
        PRIOR FILING DATE: 1998-11-13
       MUMBER OF SEQ ID NOS: 5
        SOFTWARE: PERL Program
      SEQ ID NO 4
          LENGTH: 339
          TYPE: PRT
          CRCANISM: Drosophila melanogaster
          FEATURE: -
          OTHER INFORMATION: g1794137
  US-10-025-730-4
```

65.1%;

Query Match

Score 1109; DB 15; Length 339;

```
Best Local Similarity 65.0%; Pred. No. 6.5e-93;
 Matches 217; Conservative 59; Mismatches 54; Indels 4; Gaps
                                                               3:
         4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63
           QУ
         1 MPLFGKSQKSPVELVKSLKEAINALEAGDRKVEKAQEDVSKNLVSIKNMLHGSSDAEPPA 60
Db
         64 E-AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 122
           QУ
         61 DYVVAQLSQELYNSNLLLLLIQNLHRIDFEGKKHVALIFNNLLRRQIGTRSPTVEYICTK 120
Db.
        123 PHILFMLLKGYE--APQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
            QΥ
        121 PEILFTLMAGYEDAHPEIALNSGTMLRECARYEALAKIMLHSDEFFKFFRYVEVSTFDIA 180
dU
        181 SDAFATFKDLLTRHKVLVADFLEQNYDTIF-EDYEKLLQSENYVTKRQSLKLLGELILDR 239
           QΥ
        181 SDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLLNSENYVTRRQSLKLLGELLLDR 240
Db
        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
            QУ
        241 HNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVFVANPNKPKPILDILLRNQT 300
nb
        300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK 333
QУ
            301 KLVDFLTNFHTDRSEDEQFNDEKAYLIKQIKELK 334
Do
RESULT 6
US-10-025-730-5
p Sequence 5, Application US/10025730
; Publication No. US20030045466A1
; GENERAL INFORMATION:
  APPLICANT: Tang, Y. Tom
   APPLICANT: Guegler, Karl J.
   APPLICANT: Corley, Neil C.
   APPLICANT: Gorgone, Gina A.
   TITLE OF INVENTION: CALCIUM BINDING PROTEIN
   FILE REFERENCE: PF-0635 US
  CURRENT APPLICATION NUMBER: US/10/025,730
   CURRENT FILING DATE: 2001-12-18
   PRIOR APPLICATION NUMBER: US/09/190,965
   PRIOR FILING DATE: 1998-11-13
   NUMBER OF SEQ ID NOS: 5
   SOFTWARE: PERL Program
  SEQ ID NO 5
    LENGTH: 377
    TYPE: PRT
    ORGANISM: Caenorhabditis elegans
    FEATURE: -
    OTHER INFORMATION: g1255838
 US-10-025-730-5
                      62.4%; Score 1063.5; DB 15; Length 377;
  Query Match
  Best Local Similarity 60.5%; Pred. No. 1.1e-88;
  Matches 211; Conservative 53; Mismatches 68; Indels
```

```
4 MP-LFSKSHKNPAEIVKILKDNLAILEK------QDKKTDKASEEVSKSLQAM 49
Qу
           1. MPLLFGKSHKSPADVVKTLREVLTILDKLPPPKLDKDGNIQSDKKYDKALDEVSKNVAMI 60
Db
         50 KEILCGTNEKEPPTE---AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILR 106
QУ
            61 KSFIYGNDSAEPSSEHVVQVAQLAQEVYNANILPMLIKMLPKFEFECKKDVGQIFNNLLR 120
Db
        107 RQIGTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFR 166
QУ
            121 RQIGTRSPTVEYLGARPEILIQLVQGYSVPDIALTCGLMLRESIRHDHLAKIILYSDVFY 180
Db
        167 DFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKR 226
QУ
             191 TFFLYVQSEVFDISSDAFSTFKELTTRHKAIIAEFLDSNYDTFFAQYQNLLNSKNYVTRR 240
Db
        227 QSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHK 286
Qу
            241 QSLKLLGELLLDRHNFNTMTKYISNPDNLRLMMELLRDKSPNIQYEAFHVFKVFVANPNK 300
Db
        287 TOPIVEILLKNOPKLIEFLSSFOKERTDDEOFADEKNYLIKQIRDLKKT 335
QУ
             301 PKPISDILNRNREKLVEFLSEFHNDRTDDEQFNDEKAYLIKQIQEMKSS 349
Ob
RESULT 7
US-10-029-386-32324
; Sequence 32324, Application US/10029386
 Publication No. US20030194704A1
; GENERAL INFORMATION:
  MPPLICANT: Penn, Sharron G.
  APPLICANT: Rank, David R.
  APPLICANT: Hanzel, David K.
  TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
USEFUL FOR GENE
  TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
   FILE REFERENCE: AEOMICA-X-2
   CURRENT APPLICATION NUMBER: US/10/029,386
   CURRENT FILING DATE: 2001-12-20
   NUMBER OF SEQ ID NOS: 34288
   SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
  SEQ ID NO 32324
   LENGTH: 820
    TYPE: PRT
    ORGANISM: Homo sapiens
    FEATURE:
    OTHER INFORMATION: MAP TO AC000066.1
    OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 0.87
   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
    OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.1
    OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
    OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
    OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.3
    OTHER INFORMATION: SWISSPROT HIT: Q99996, EVALUE 0.00e+00
US-10-029-386-32324
```

```
Best Local Similarity 20.1%; Pred. No. 0.0072;
 Matches 77; Conservative 75; Mismatches 116; Indels 115; Gaps 15;
        18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77
Qу
           358 IEKLKDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE---- 404
         78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI------GTRSPTVEYISAHPHI 125
Qv
            405 --ISKLKDLQQSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 460
Db
        126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
Qv
                     | :| |
        461 LEKQMKEKE----
                                       ----NDLQEKFAÇLEAEN-SILKDEKK 491
        186 TFKDLLTRH------KVLVADFLE-QNYDT1FEDYEKLLQSENYVTKRQSLKLLGELIL 237
ÕУ
           492 TLEDMLKIHTPVSQEERLIFLDSIKSKSKDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 551
Db
        238 DRHNFAIMTK-----
                                           ----YISKPENLKLMMNLLRD 264
OV
                                               1: |: |
        552 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNKQELEYKSKLKALNEELHLQRI 611
Db
        255 KSPNIQFEA--FHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320
Ov
            612 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVVEKDTTELMEKLEVTKREKLELSQRLSDL 668
Dh
        321 -----EKNYLIKQIRDLKK 334
QУ
                  | | ::| :::: ||:
        669 SEQLKOKHGEISFLNEEVKSLKQ 691
Db'
RESULT 3
US-10-080-608A-11
; Sequence 11, Application US/10080608A
; Publication No. US20030198956A1
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Faul
; APPLICANT: Williams, Mark
  TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/10/080,608A
; GURRENT FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 11
   LENGTH: 3878
   TYPE: PRT
  ORGANISM: Homo sapiens
US-10-080-608A-11
                      7.5%; Score 128.5; DB 12; Length 3878;
  Query Match
  Best Local Similarity 20.1%; Pred. No. 0.065;
  Matches 77; Conservative 75; Mismatches 116; Indels 115; Gaps
       18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77
```

```
664 IEKLKDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE---- 710
Db
         78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI------GTRSPTVEYISAHPHI 125
             :::||| |::::| || |:::| || |:::|
QУ
        711 --ISKLKDLQQSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 766
Db
         126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
                                           | :: | :| |
QΥ
            | :| |
                                          ----NDLQEKFAQLEAEN-SILKDEKK 797
         767 LEKQMKEKE-----
         186 TFKDLLTRH------KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237
            Qу
         798 TLEDMLKIHTPVSQEERLIFLDSIKSKSKDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 857
αd
                                              ----YISKPENLKLMMNLLRD 264
         238 DRHNFAIMTK-----
                                                   1 | | : | | : : | |
ÇУ
            |: |: |
         858 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNKQELEYKSKLKALNEELHLQRI 917
Db.
         265 KSPNIQFEA--FHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320
            QУ
         918 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVVEKDTTELMEKLEVTKREKLELSQRLSDL 974
. כוֹּה
         321 -----EKNYLIKQIRDLKK 334
              | ::| :::: ||:
         975 SEQLKQKHGEISFLNEEVKSLKQ 997
 RESULT 9
 US-10-171-311-4
 ; Sequence 4, Application US/10171311
 ; Publication No. US20030087270A1
 ; GENERAL INFORMATION:
   APPLICANT: Schlegel, Robert
   APPLICANT: Chen, Yan
   APPLICANT: Zhao, Xumei
   APPLICANT: Monahan, John
   APPLICANT: Kamatkar, Shubhangi
   APPLICANT: Glatt, Karen
   APPLICANT: Gannavarapu, Manjula
APPLICANT: Hoersh, Sebastian
   TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
    TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
    TITLE OF INVENTION: OF CERVICAL CANCER
    FILE REFERENCE: MRI-035
   CURRENT APPLICATION NUMBER: US/10/171,311
    CURRENT FILING DATE: 2002-06-12
    PRIOR APPLICATION NUMBER: US 60/298,159
    PRIOR FILING DATE: 2001-06-13
    PRIOR APPLICATION NUMBER: US 60/298,155
    PRIOR FILING DATE: 2001-06-13
    PRIOR APPLICATION NUMBER: US 60/335,936
   PRIOR FILING DATE: 2001-11-14
    NUMBER OF SEQ ID NOS: 238
    SOFTWARE: FastSEQ for Windows Version 4.0
   SEQ ID NO 4
    LENGTH: 3899
```

```
TYPE: PRT
   ORGANISM: Homo sapiens
US-10-171-311-4
                       7.5%; Score 128.5; DB 15; Length 3899;
 Ouery Match
 Best Local Similarity 20.1%; Pred. No. 0.066;
         77; Conservative 75; Mismatches 116; Indels 115; Gaps 15;
         18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77
            ÒЯ
        652 IEKLKDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE---- 698
Dh
         78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI------GTRSPTVEYISAHPHI 125
             QУ
         699 --ISKLKDLQQSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 754
Db
         126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185.
QУ
                                            1 :: | : | |
               : | | .
                              -----NDLQEKFAQLEAEN-SILKDEKK 785
         755 LEKQMKEKE--
Db
         186 TFKDLLTRH-----KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237
            QУ
         786 TLEDMLKIHTPVSQEERLIFLDSIKSKSKDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 845
DΒ
                                             ----YISKPENLKLMMNLLRD 264
         238 DRHNFAIMTK---
QУ
                                                  | | | : | :: | |
             1: 1: 1:
         846 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNKQELEYKSKLKALNEELHLQRI 905
Db
         ·265 KSPNIQFEA--FHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320
            QУ
         906 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVVEKDTTELMEKLEVTKREKLELSQRLSDL 962
         321 -----EKNYLIKQIRDLKK 334
QY
                    | ::| :::: ||:
         963 SEQLKQKHGEISFLNEEVKSLKQ 985
 Db
 RESULT 10
 US-10-171-311-2
 ; Sequence 2, Application US/10171311
 ; Publication No. US20030087270A1
 GENERAL INFORMATION:
   APPLICANT: Schlegel, Robert
   APPLICANT: Chen, Yan
   APPLICANT: Zhao, Xumei
   APPLICANT: Monahan, John
              Kamatkar, Shubhangi
   APPLICANT:
   APPLICANT: Glatt, Karen
   APPLICANT: Gannavarapu, Manjula
   APPLICANT: Hoersh, Sebastian
   TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
   TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
   TITLE OF INVENTION: OF CERVICAL CANCER
   FILE REFERENCE: MRI-035
   CURRENT APPLICATION NUMBER: US/10/171,311
   CURRENT FILING DATE: 2002-06-12
    PRIOR APPLICATION NUMBER: US 60/298,159
```

```
PRIOR FILING DATE: 2001-06-13
    PRIOR APPLICATION NUMBER: US 60/298,155
    PRIOR FILING DATE: 2001-06-13
    PRIOR APPLICATION NUMBER: US 60/335,936
    PRIOR FILING DATE: 2001-11-14
    NUMBER OF SEQ ID NOS: 238
     SOFTWARE: FastSEQ for Windows Version 4.0
   SEQ ID NO 2
      LENGTH: 3907
       TYPE: PRT
      ORGANISM: Homo sapiens
US-10-171-311-2
                                                   7.5%; Score 128.5; DB 15; Length 3907;
   Query Match
   Best Local Similarity 20.1%; Pred. No. 0.066;
   Matches 77: Conservative 75; Mismatches 116; Indels 115; Gaps
                     18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77
Qу
                                                                                                                         li l
                           |||:
                   652 IEKLKDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE---- 698
Db
                     78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI------GTRSPTVEYISAHPHI 125
                              :::||| |::::| || || |:::|
QУ
                    699 --ISKLKDLQQSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 754
Db
                    126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
 \Omega V
                                                                                                   ---NDLQEKFAQLEAEN-SILKDEKK 785
                    755 LEKQMKEKE-----
 づし
                    186 TFKDLLTRH------KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237.
                           THE THE RESERVE OF THE PERSON 
 QУ
                    786 TLEDMLKIHTPVSQEERLIFLDSIKSKSKDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 845
 Db
                                                                                                        ---YISKPENLKLMMNLLRD 264
                    238 DRHNFAIMTK-
 Q_{i}^{*}
                                                                                                                1 || : | ::| |
                              1: 1:
                    846 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNKQELEYKSKLKALNEELHLQRI 905
 dC
                    265 KSPNIQFEA--FHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320
                                    QY
                     906 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVVEKDTTELMEKLEVTKREKLELSQRLSDL 962
 On
                     321 -----EKNYLIKQIRDLKK 334
 ŨУ
                                              | ::| :::: ||:
                     963 SEQLKQKHGEISFLNEEVKSLKQ 985
  RESULT 11
  US-10-370-685-100
  ; Sequence 100, Application US/10370685
  ; Publication No. US20030215903A1
  ; CENERAL INFORMATION:
      APPLICANT: Hyman, Paul
      APPLICANT: Goldberg, Edward
       TITLE OF INVENTION: Nanostructures Containing PNA Joining and Functional
  Elements
      FILE REFERENCE: NANF.P-004
        CURRENT APPLICATION NUMBER: US/10/370,685
```

```
CURRENT FILING DATE: 2003-02-21
 PRIOR APPLICATION NUMBER: 10/080,608
  PRIOR FILING DATE: 2002-02-21
  NUMBER OF SEQ ID NOS: 159
  SOFTWARE: PatentIn version 3.2
 SEO ID NO 100
   LENGTH: 3911
   TYPE: PRT
   ORGANISM: human
US-10-370-685-100
                      7.5%; Score 128.5; DB 12; Length 3911;
 Ouery Match
 Best Local Similarity 20.1%; Pred. No. 0.066;
 Matches 77; Conservative 75; Mismatches 116; Indels 115; Gaps 15;
        18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77
           ŨУ
        664 IEKLKDNIGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE---- 710
da
         78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI------GTRSPTVEYISAHPHI 125
             :::||| |::::| || ||:::|
Öλ
         711 --ISKLKDLQQSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 766
         126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
                                       QУ
           -----NDLQEKFAQLEAEN-SILKDEKK 797
         757 LEKOMKEKE-----
         106 TFKDLLTRH------KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 2378
            QΥ.
         798 TLEDMLKIHTPVSQEERLIFLDSIKSKSKDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 857
                                               ---YISKPENLKLMMNLLRD 264
         238 DRHNFAIMTK-
QY
                                                 |: |: |
         858 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNKQELEYKSKLKALNEELHLQRI 917
Db.
         265 KSPNIQFEA--FHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320
                QY
         918 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVVEKDTTELMEKLEVTKREKLELSQRLSDL 974
 Db
         321 -----EKNYLIKQIRDLKK 334
 QУ
                    1 ::| :::: ||:
         975 SEQLKQKHGEISFLNEEVKSLKQ 997
 Db
 RESULT 12
 US-10-171-311-8
 , Sequence 8, Application US/10171311
 ; Publication No. US20030087270A1
 ; GENERAL INFORMATION:
   APPLICANT: Schlegel, Robert
   APPLICANT: Chen, Yan
    APPLICANT: Zhao, Xumei
    APPLICANT: Monahan, John
 ; APPLICANT: Kamatkar, Shubhangi
 ; APPLICANT: Glatt, Karen
 , APPLICANT: Gannavarapu, Manjula
, APPLICANT: Hoersh, Sebastian
```

```
TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
   TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
   TITLE OF INVENTION: OF CERVICAL CANCER
    FILE REFERENCE: MRI-035
    CURRENT APPLICATION NUMBER: US/10/171,311
    CURRENT FILING DATE: 2002-06-12
    PRIOR APPLICATION NUMBER: US 60/298,159
    PRIOR FILING DATE: 2001-06-13
    PRIOR APPLICATION NUMBER: US 60/298,155
    PRIOR FILING DATE: 2001-06-13
    PRIOR APPLICATION NUMBER: US 60/335,936
    PRIOR FILING DATE: 2001-11-14
    NUMBER OF SEQ ID NOS: 238
    SOFTWARE: FastSEQ for Windows Version 4.0
   SEQ ID NO 9
      LENGTH: 3917
      TYPE: PRT
  ORGANISM: Homo sapiens
US-10-171-311-8
                                                   7.5%; Score 128.5; DB 15; Length 3917;
   Query Match
   Best Local Similarity 20.1%; Pred. No. 0.066;
   Matches 77; Conservative 75; Mismatches 116; Indels 115; Gaps
                     18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77
QУ
                                                                                                                          698 S52 TEKLKDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE---- 698
Db
                     78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI-----GTRSPTVEYISAHPHI 125
                           Sand Hill Indiana All Marie Control in i
Ç٧.
                   699 -- ISKLKDLQQSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 754
DO.
                    126 LFMLLKGYEAPOIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
                                                                                                 QУ
                            1 : 1
                                                                                          ----NDLQEKFAQLEAEN-SILKDEKK 785
                    755 LEKQMKEKE-----
 Dlo
                    186 TFKDLLTRH------KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237
                            1 4 4 4 7 7 7 7 8 8 8 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4 5 1 4
 QУ
                    786 TLEDMLKIHTPVSQEERLIFLDSIKSKSKDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 845
                                                                                                            --YISKPENLKLMMNLLRD 264
                    238 DRHNFAIMTK-----
                                                                                                               1 | | : | :: | |
 QУ
                              1: |: |.
                    846 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNKQELEYKSKLKALNEELHLQRI 905
 Db
                     265 KSPNIQFEA--FHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320
                                    Qу
                     906 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVVEKDTTELMEKLEVTKREKLELSQRLSDL 962
  Db
                     321 -----EKNYLIKQIRDLKK 334
  ÇV
                                              | ::| :::: ||:
                     963 SEQLKQKHGEISFLNEEVKSLKQ 985
  Db
```

RESULT 13

US-10-171-311-6

; Sequence 6, Application US/10171311 ; Publication No. US20030087270A1

```
GENERAL INFORMATION:
 APPLICANT: Schlegel, Robert
 APPLICANT: Chen, Yan
APPLICANT: Zhao, Xumei
 APPLICANT: Monahan, John
 APPLICANT: Kamatkar, Shubhangi
 APPLICANT: Glatt, Karen
 APPLICANT: Gannavarapu, Manjula
 APPLICANT: Hoersh, Sebastian
 TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND METHODS FOR
 TITLE OF INVENTION: IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
 TITLE OF INVENTION: OF CERVICAL CANCER
 FILE REFERENCE: MRI-035
 CURRENT APPLICATION NUMBER: US/10/171,311
 CURRENT FILING DATE: 2002-06-12
 PRIOR APPLICATION NUMBER: US 60/298,159
 PRIOR FILING DATE: 2001-06-13
 PRIOR APPLICATION NUMBER: US 60/298,155
 PRIOR FILING DATE: 2001-06-13
 PRIOR APPLICATION NUMBER: US 60/335,936
 PRIOR FILING DATE: 2001-11-14
 NUMBER OF SEQ ID NOS: 238
 SOFTWARE: FastSEQ for Windows Version 4.0
 SEO ID NO 6
  LENGTH: 3925
   TYPE: PRT
 ORGANISM: Homo sapiens
ys-10-171-311-6
                       7.5%; Score 128.5; DB 15; Length 3925; The
 Query Match
 Rest Local Similarity 20.1%; Pred. No. 0.067;
 Statistiches 77; Conservative 75; Mismatches 116; Indels 115; Gaps
         18 VKILKDNLATLEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77
            652 IEKLKDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE---- 698
Db
         78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI------GTRSPTVEYISAHPHI 125
            \Omega_{\lambda}
        699 -- ISKLKDLQQSLVNSKSEEMTLQI- NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 754
nh
        126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
QУ
                                       1 : 1
                                  ------NDLQEKFAQLEAEN-SILKDEKK 785
         755 LEKOMKEKE----
        186 TFKDLLTRH------KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237
            786 TLEDMLKIHTPVSQEERLIFLDSIKSKSKDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 845
Db
                                                 -YISKPENLKLMMNLLRD 264
         238 DRHNFAIMTK------
                                                  | | | : | :: | |
             1: 1: 1
         346 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNKQELEYKSKLKALNEELHLQRI 905
Db
         265 KSPNIQFEA--FHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320
Oν
               906 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVVEKDTTELMEKLEVTKREKLELSQRLSDL 962
Db
```

```
321 -----EKNYLIKQIRDLKK 334
Qу
                      | ::| :::: ||:
         963 SEQLKQKHGEISFLNEEVKSLKQ 985
dd
RESULT 14
US-09-864-761-47959
; Sequence 47959, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
  APPLICANT: Penn, Sharron G.
   APPLICANT: Rank, David R.
   APPLICANT: Hanzel, David K.
   APPLICANT: Chen, Wensheng
   TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES
USEFUL FOR
   TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
   FILE REFERENCE: Aeomica-X-1
   CURRENT APPLICATION NUMBER: US/09/864,761
   CURRENT FILING DATE: 2001-05-23
   PRIOR APPLICATION NUMBER: US 60/180,312
   PRIOR FILING DATE: 2000-02-04
   PRIOR APPLICATION NUMBER: US 60/207,456
   PRIOR FILING DATE: 2000-05-26
   PRIOR APPLICATION NUMBER: US 09/632,366
   PRIOR FILING DATE: 2000-08-03
   TRIOR APPLICATION NUMBER: GB 24263.6
   URIOR FILING DATE: 2000-10-04
   PRIOR APPLICATION NUMBER: US 60/236,359
   PRIOR FILING DATE: 2000-09-27
   PRIOR APPLICATION NUMBER: PCT/US01/00666
 ; PRIOR FILING DATE: 2001-01-30
   PRIOR APPLICATION NUMBER: PCT/US01/00667
    FRIOR FILING DATE: 2001-01-30
   PRIOR APPLICATION NUMBER: PCT/US01/00664
   PRIOR FILING DATE: 2001-01-30
    PRIOR APPLICATION NUMBER: PCT/US01/00669
    PRIOR FILING DATE: 2001-01-30
    PRIOR APPLICATION NUMBER: PCT/US01/00665
    PRIOR FILING DATE: 2001-01-30
   PRIOR APPLICATION NUMBER: PCT/US01/00668
   PRIOR FILING DATE: 2001-01-30
 ; PRIOR APPLICATION NUMBER: PCT/US01/00663
    PRIOR FILING DATE: 2001-01-30
   PRIOR APPLICATION NUMBER: PCT/US01/00662
    PRIOR FILING DATE: 2001-01-30
   PRIOR APPLICATION NUMBER: PCT/US01/00661
    PRIOR FILING DATE: 2001-01-30
    PRIOR APPLICATION NUMBER: PCT/US01/00670
    PRIOR FILING DATE: 2001-01-30
    PRIOR APPLICATION NUMBER: US 60/234,687
    PRIOR FILING DATE: 2000-09-21
    PRIOR APPLICATION NUMBER: US 09/608,408
    PRIOR FILING DATE: 2000-06-30
    PRIOR APPLICATION NUMBER: US 09/774,203
```

PRIOR FILING DATE: 2001-01-29 NUMBER OF SEQ ID NOS: 49117

10

```
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
 SEQ ID NO 47959
      LENGTH: 660
      TYPE: PRT
      ORGANISM: Homo sapiens
     FEATURE:
     OTHER INFORMATION: MAP TO AJ010770.1
     OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL =
     OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.1
      OTHER INFORMATION: SWISSPROT HIT: Q99323, EVALUE 3.00e-17
      OTHER INFORMATION: EST_HUMAN HIT: AU132932.1, EVALUE 1.00e-105
US-09-864-761-47959
                                              7.3%; Score 125; DB 9; Length 660;
   Best Local Similarity 20.5%; Pred. No. 0.011;
   Query Match
   Matches 75; Conservative 68; Mismatches 119; Indels 104;
                   18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77
                         342 IEKLKDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE---- 388
Db
                    78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI------GTRSPTVEYISAHPHI 125
                          QУ
                  389 -- ISKLKDLQQSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 444
D.o
                125 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
                      \Omega Y
                                                                        -----NDLQEKFAQLEAEN-SILKDEKK 475
                   445 LEKOMKEKE-----
 dŒ
                  186 TFKDLLTRH------KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237
                           Palaton Control of the first terms of the fall of the first terms of t
 Çv
                   476 TLEDMLKIHTPVSQEERLIFLDSIKSKSKDSVWEKEIEILIEENEDLKQQCIQLNEEIEK 535
                                                                                                    ---YISKPENLKLMMNLLRD 264
                   238 DRHNFAIMTK-----
                                                                                                         | || : | ::| |
 QУ
                            1: 1: 1
                   536 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNKQELEYKSKLKALNEELHLQRI 595
                    265 KSPNIQFEA--FHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEK 322
                                  Qу
                    596 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVVEKDTTELMEKLEVTKREKLELSQRLSDL 652
 Dο
                    323 NYLIKQ 328
  QV:
                           : :||
                    653 SEQLKQ 658
  Db
  RESULT 15
  US-10-023-634-18
  ; Sequence 18, Application US/10023634
   ; Publication No. US20030236389A1
   ; GENERAL INFORMATION:
      APPLICANT: Shimkets, Richard A
       APPLICANT: Colman, Steven D
       APPLICANT: Spytek, Kimberly A
       APPLICANT: Ballinger, Robert A
   ; APPLICANT: Guo, Xiaojia
   ; APPLICANT: Tchernev, Velizar T
```

```
Shenoy, Suresh G
APPLICANT:
           Li, Li
APPLICANT:
           Ellerman, Karen
APPLICANT:
            Zerhusen, Bryan D
APPLICANT:
           Patturajan, Meera
APPLICANT:
APPLICANT: Casman, Stacie J
APPLICANT: Boldog, Ferenc
APPLICANT: Gusev, Vladimir Y
            Burgess, Catherine E
APPLICANT:
APPLICANT: Edinger, Shlomit R
            Gangolli, Esha A
APPLICANT:
            Malyankar, Uriel M
APPLICANT:
            Gunther, Erik
 APPLICANT:
            Smithson, Glennda
 APPLICANT:
 APPLICANT: Millet, Isabelle
 APPLICANT: Gerlach, Valerie
 TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of
 TITLE OF INVENTION: Using the Same
 FILE REFERENCE: 21402-221
 CURRENT APPLICATION NUMBER: US/10/023,634
 CURRENT FILING DATE: 2002-06-28
 PRIOR APPLICATION NUMBER: 60/256,025
 PRIOR FILING DATE: 2000-12-15
 PRIOR APPLICATION NUMBER: 6C/265,163
 WRIOR FILING DATE: 2001-01-30
 URIOR APPLICATION NUMBER: 60/272,929
 PRIOR FILING DATE: 2001-03-02
 PRIOR APPLICATION NUMBER: 60/274,864
 PRIOR FILING DATE: 2001-03-09
 PRIOR APPLICATION NUMBER: 60/276,688
 PRIOR FILING DATE: 2001-03-16
 PRIOR APPLICATION NUMBER: 60/277,880
 PRIOR FILING DATE: 2001-03-22
 PRIOR APPLICATION NUMBER: 60/286,409
  PRIOR FILING DATE: 2001-04-25
  PRIOR APPLICATION NUMBER: 60/309,246
 PRIOR FILING DATE: 2001-07-31
 PRIOR APPLICATION NUMBER: 60/315,600
 PRIOR FILING DATE: 2001-08-29
 NUMBER OF SEQ ID NOS: 132
  SOFTWARE: PatentIn Ver. 2.1
 SEO ID NO 78
   LENGTH: 709
   TYPE: PRT
   ORGANISM: Homo sapiens
U3-10-023-634-18
                         6.8%; Score 116.5; DB 12; Length 709;
 Query Match
 Best Local Similarity 19.6%; Pred. No. 0.073;
 Matches 79; Conservative 64; Mismatches 146; Indels 115; Gaps
                                                                         12;
           3 KMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPP 62
             QУ
         179 KLQVTQRSLEESQGKIAQLEGKLVSIEKE--KIDEKS-ETEKLLEYIEEISCASDQVEKY 235
Db
          63 TEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 122
Qу
                                          : | |
                 : | | | : |
```

Db	236	KLDIAQLEENLKEKNDEILSLKQSLEENIVILSKQVE 2/2	
Qу		PHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRD 167 : ::   ::     : :    : :    : :    : :    : :    : :    : :    : :      : :    : :    : :    : :    : :    : :  :	
Db	273		
Qу		FFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLL 217 :   : :     : :     : :     : :     : :       : : :         : : :         : : :         : : :         : : :       : : :         : : :       : : :       : : :       : : :       : : :       : : :       : : :       : : :       : : :         : : :       : : :       : : :       : : :       : : :       : : :       : : :       : : :       : : :       : : :       : : :       : : :       : : :       : : :       : : :       : : :       : : :       : : :       : : :         : : :         : : :         : : :         : : :         : : :         : : :         : : :         : : :         : : : :         : : :         : : :       : : :         : : :         : : : :         : : : :       : : :       : : :       : : :       : : :       : : :       : : :	
Db		KLQQKELQIDSLLQQEKELSSSLHQKLCSFQEEMVKEKNLFEEELKQTLDELDKLQQKEE 378	
Qy		QSENYVTKRQSLKLLGELILDRHNFAIMTKY 248	
Db		QAERLVKQLEEEAKSRAEELKLLEEKLKGKEAELEKSSAAHIQAIBBBQBKIBSHVQSSS	
Qу		ISKPENLKLMMNLLRPKSPNIQFEAFHVFKVFVASPHKTQPIVEILL 295	
Db	439	:  :	
QY		KNQPKLIEFLSSFQKERTD-DEQFADEKNYLIKQIRD 331	
Db	499	DLQTKSALKETEIKEITVSFLQKITDLQNQLKQQEEDFRKQLED 542	

Search completed: January 7, 2004, 16:52:26
Job time: 37 secs

# GenCore version 5.1.6 Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

January 7, 2004, 16:44:17; Search time 41 Seconds Run on:

(without alignments)

2121.067 Million cell updates/sec

US-10-088-872-2 Title:

1704

Perfect score: 1 MKKMPLFSKSHKNPAEIVKI......FADEKNYLIKQIRDLKKTAP 337 Sequence:

BLOSUM62 Scoring table:

Gapop 10.0 , Gapext 0.5

830525 seqs, 258052604 residues Searched:

Total number of hits satisfying chosen paraméters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Fost-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

SPTREMBL 23:\* - abase :

1: sp\_archea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriap:\*

17: sp\_archeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

### SUMMARIES

왕

Query Result Score Match Length DB ID No.

Description

						•
1	1684	98.8	337	11	Q8BG52	Q8bg52 mus musculu
2	1669	97.9	334	11	Q91WB8	Q91wb8 mus musculu
3	1663	97.6	334	11	Q91YL0	Q91yl0 mus musculu
4	1462	85.8	289	4	Q96FG1	Q96fg1 homo sapien
5	1381	81.0	341	11	Q8VDZ8	Q8vdz8 mus musculu
6.	1066.5	62.6	636	5	Q21643	Q21643 caenorhabdi
7	875	51.3	205	11	Q8K312	Q8k312 mus musculu
3	709.5	41.6	333	10	Q8H5L9	Q8h5l9 oryza sativ
Э	671.5	39.4	345	10	Q8L9L9	Q81919 arabidopsis
10	590	34.6	322	10	Q8LIF3	Q8lif3 oryza sativ
11	435	25.5	103	11	Q8K038	Q8k038 mus musculu
12	134.5	7.9	677	16	025188	O25188 helicobacte
13	128	7.5	430	1.6	026049	026049 helicobacte
14	123.5	7.2	1285	16	Q9WXU3	Q9wxu3 thermotoga
15	120	7.0	1175	17	Q58914	Q58914 methanococc
16	119.5	7.0	1056	16	Q8REF7	Q8ref7 fusobacteri
17	119	7.0	1111	5	Q9VGE4	Q9vge4 drosophila
18	118.5	7.0	554	5	Q8IN90	Q8in90 drosophila
19	118.5	7.0	670	5	Q9VEC7	Q9vec7 drosophila
20	118.5	7.0	670	5	Q9NFM7	Q9nfm7 drosophila
211	117	6.9	808	5	Q8T133.	Q8t133 dictyosteli
22	117	6.9	808	5	Q9GSH4	Q9gsh4 dictyosteli
23	116.5	6.8	1135	5	Q9NJQ4	Q9njq4 paramecium 👉
24	116	6.8	911	16	Q8EUI7	Q8eui7 mycoplasma
25	116	6.8	1389	5	Q8I293	Q8i293 plasmodium
25	. 1.15.5	6.8	1111	5.	Q9U0K5	Q9u0k5 plasmodium
.27	1.4 . 5	6 : ទិ	1946	5	097291	O97291 plasmodium
28	2.15	6.7	473	11	Q8R436	Q8r436 mus musculu
29	. 3.15	6.7	2518	5	Q8IEH2	Q8ieh2 plasmodium
30	144:5	6.7	1941	5	Q8IAK6	Q8iak6 plasmodium
31	$\pm 14$	5.7	743	13	Q9YGE7	Q9yge7 oncornynchu
32	113.5	6.7	833	4	Q9UF54	Q9uf54 homo sapien
30	213.5	6.7	951	5	Q5/AECe	Q9vec6 drosophila
34	<b>3.13.5</b>	6.7	984	5 .	Q8IN89	Q8in89 drosophila
.35	15.3	, 6.6	474	5 '	097233	O97233 plasmodium
36	113	6.6	647	. 11	-	Q8ca10 mus musculu
37	111.5	6.5	1925	5	Q8I2D1	Q8i2d1 plasmodium
3.0	111.5	6.5	2429	5	Q9VFB1	Q9vfb1 drosophila.
39	2.1.1.5	6.5	2771	5	Q26216	Q26216 plasmodium
4()	111	6.5	2166	16	051465	051465 borrelia bu
41	211	6.5	2819	16	Q98QP8	Q98qp8 mycoplasma
42	1110	5.5	461		077390	077390 plasmodium
43	110	6.5	1183	2	086064	086064 helicobacte
44	110	6.5	1758	5	Q8I1K5	Q8i1k5 plasmodium
45	109.5	6.4	457	16	Q9PQM0	Q9pqm0 ureaplasma
	•					

## ALIGNMENTS

# PESULT 1 Q8BG52 ID Q8BG52; PRELIMINARY; PRT; 337 AA. AC Q8BG52; DT 01-MAR-2003 (TrEMBLrel. 23, Created) DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update) DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)

```
MO25-like protein homolog.
DE
    Mus musculus (Mouse).
OS
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC
    NCBI TaxID=10090;
OX
RM
    SEOUENCE FROM N.A.
RP
    STRAIN=C57BL/6J; TISSUE=Eye, Pituitary, and Testis;
RC
    MEDLINE=22354683; PubMed=12466851;
RX
    The FANTOM Consortium,
RA
    the RIKEN Genome Exploration Research Group Phase I & II Team;
RA
    "Analysis of the mouse transcriptome based on functional annotation of
RT
    60,770 full-length cDNAs.";
\Sigma\Sigma
    Nature 420:563-573 (2002).
PI.
    EMBL: AK030474; BAC26978.1; -.
DR
    EMBL; AK053642; BAC35457.1; -.
DR
    EMBL; AK076758; BAC36470.1; -.
DR
              337 AA; 39105 MW; C62B5B58095A98C8 CRC64;
    SEOUENCE
SQ
                              Score 1684; DB 11; Length 337;
                       98.8%;
  Query Match
                             Pred. No. 1.1e-110;
                       98.5%;
  Best Local Similarity
                                                Indels
                                                            Gaps
                                Mismatches
  Matches 332; Conservative
                             2;
           1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKE 60
Qy
            ! MKKMPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNDKE 60
מנו
          61 PPTEAVAQLAQELYSSGLLVTLIADLQLTDFEGKKDVTQIFNNILRRQIGTRSPTVEYIS 120
OV
             51 PPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRCPTVEYIS 120
Db
         121 AHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
QΥ
             121 SHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
Db
         181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
QУ
             181 SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRH 240
Db
         241 NFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPK 300.
Qу
             241 NFTIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPK 300
Db
         301 LIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
 ÒУ
             301 LIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKAAP 337
 Db
 RESULT 2
 091WB8
                                      334 AA.
                               PRT:
                PRELIMINARY;
 ΙĿ
     091WB8
 AC
     091WB8;
     01-DEC-2001 (TrEMBLrel. 19, Created)
 DT
     01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT
     01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DT
     Similar to hypothetical protein FLJ12577 (MO25-like protein
 DE
 DE
     homolog).
```

```
OS
     Mus musculus (Mouse).
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC.
OC
     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX
     NCBI TaxID=10090;
RN
     [1]
     SEQUENCE FROM N.A.
RP
RC
     TISSUE=Salivary gland;
     Strausberg R.;
RA
     Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RL
RN
RP
     SEOUENCE FROM N.A.
RC.
     STRAIN=C57BL/6J; TISSUE=Testis;
RX
     MEDLINE=22354683; PubMed=12466851;
RA
     The FANTOM Consortium,
\mathbb{R}A
     the RIKEN Genome Exploration Research Group Phase I & II Team;
     "Analysis of the mouse transcriptome based on functional annotation of
RT
     60,770 full-length cDNAs.";
RT
RI.
     Nature 420:563-573(2002).
 DR.
     EMBL; BC015128; AAH16128.1; -.
     EMBL; AK076867; BAC36513.1; -.
DR
DR
     InterPro; IPR004892; Mo25.
     Pfam; PF03204; Mo25; 1.
DR
ΚM
     Hypothetical protein.
SO
     SEQUENCE
             334 AA; 38718 MW;
                                822F04A87FB4EB6F CRC64;
                       97.9%; Score 1669; DB 11;
  Querry Match
                                                Length 334;
                       98.5%; Pred. No. 1.3e-109;
  Best Local Bimilarity
  Matches 329; Conservative
                             2; Mismatches
                                             3;
                                                Indels
                                                             Gaps
           4 MPI@SKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63
\Omega_{\mathcal{F}}
             1 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTMDKEPPT 60
1.0
          64 EAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHP 123:
, KB
             51 EAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRCPTVEYISSHP 120
Db
         124 HILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDA 183 😁
QY
             121 HILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDA 180-
(C)
         194 FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFA 243 🤫
Qy
             Db
         181 FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFT 240 ::
         # IMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIE 303
QУ
             241 IMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIE 300
Dр
         304 FLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
QΫ
             301 FLSSFOKERTDDEQFADEKNYLIKQIRDLKKAAP 334
Dh
```

RESULT 3 Q91YL0

```
01-DEC-2001 (TrEMBLrel. 19, Created)
DT
    01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT
    01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DT
    Similar to hypothetical protein FLJ12577.
DE
    Mus musculus (Mouse).
os
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC.
    NCBI TaxID=10090;
OX
    [1]
RN
    SEQUENCE FROM N.A.
RP
    Strausberg R.;
RA
    Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RL
    EMBL; BC016546; AAH16546.1; -.
DR
    InterPro; IPR004892; Mo25.
DR
    Pfam; PF03204; Mo25; 1.
DP.
    Hypothetical protein.
KW
                               5F9765360653750E CRC64;
                     38761 MW;
              334 AA:
    SEOUENCE
SQ
                             Score 1663; DB 11;
                                                Length 334;
                       97.6%;
  Ouery Match
                             Pred. No. 3.3e-109;
                      98.2%;
  Best Local Similarity
                                                Indels
                                                            Gaps
                                                                   0;
                             2; Mismatches
         328; Conservative
  Matches
          4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63
QΥ
            MPLFSKSHKNPAEIVKILKDNLAILEKODKKTDKASEEVSKSLQAMKEILCGTNDKEPPT 60
         34 EAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHP 12334
            51 EAVAQLAQELYSSGLLVTLIADEQLIDFEGKKDVTQIFNNILRRQIGTPCPTVEYISSHP 120-
         134 HILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDA 183
<u>(</u>
             321 HILPMLLKGYEAPQIALRCGIMLRECIRHEPLAKTILFSNQFRDFFKYVELSTFDIASDA 180
Db
         184 FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFA 243
Qy
             131 FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLRGELILDRHNFT 240
TO
         244 IMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIE 303
ΩУ
             241 IMTKYTSKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIE 300
ρþ
         30% FLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
 Qy
             301 FLSSFQKERTDDEQFADEKNYLIKQIRDLKKAAP 334
 Dp
 RESULT 4
 096FG1
                               PRT:
                                     289 AA.
                PRELIMINARY;
     Q96FC1
 AC
     01-DEC-2001 (TrEMBLrel. 19, Created)
 T)T
     01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT
     01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DT
     Hypothetical protein.
 DE:
     Homo sapiens (Human).
 OS
```

091YL0;

AC

```
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OC
    NCBI TaxID=9606;
XO
RN
    [1]
    SEOUENCE FROM N.A.
RP
    TISSUE=Placenta;
RC
    Strausberg R.;
RA
    Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
RL
    EMBL; BC010993; AAH10993.1; -.
DR
    InterPro; IPR004892; Mo25.
DR
    Pfam; PF03204; Mo25; 1.
DR
    Hypothetical protein.
ΚW
             289 AA; 33738 MW; F57B9EFCF6ABF2D7 CRC64;
SO
    SEQUENCE
                       85.8%; Score 1462; DB 4; Length 289;
 Query Match
                       99.7%; Pred. No. 3.8e-95;
 Best Local Similarity
                                                                    0:
                                                Indels
                                                             Gaps
 Matches 288; Conservative
                             0; Mismatches
                                             1;
         49 MKEILCGTNEKEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQ 108
Qy
            1 MKEILCGTNEKEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEEKKDVTQIFNNILRRQ 60
Db
         109 IGTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDF 168.
QУ
            61 IGTESPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDF 120
         169 FKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQS 228
QУ
            180. BKTVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQS 180.
275
         239 LKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQ 288
            181 LKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQ 2408
Db
         289 PIVEILLKNOPKLIEFLSSFOKERTDDEOFADEKNYLIKQIRDLKKTAP 337
QУ
            241 PIVEILLKNOPKLIEFLSSFOKERTDDEOFADEKNYLIKOIRDLKKTAP 289
Db
RESULT 5
Q3VDZ8
               PRELIMINARY;
                             PRT;
                                     341 AA.
    Q8VDZ8
.ID
    Q8VDZ8;
AC
    01-MAR-2002 (TrEMBLrel. 20, Created)
DT
    01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
D\mathcal{R}
    01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DT
    MO25 protein.
DE
GN
    CAB39.
    Mus musculus (Mouse).
OS
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CC
    NCBI TaxID=10090;
OX
     [1]
RM
    SEQUENCE FROM N.A.
RΡ
    Strausberg R.;
RA
    Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
RЪ
    EMBL; BC020041; AAH20041.1; -.
DR
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC

```
MGD; MGI:107438; Cab39.
DR
    InterPro: IPR004892; Mo25.
את
    Pfam; PF03204; Mo25; 1.
DR
            341 AA; 39843 MW; E7FECA529D6FE811 CRC64;
SO
    SEOUENCE
                      81.0%; Score 1381; DB 11;
                                               Length 341;
 Query Match
                            Pred. No. 2.3e-89;
 Best Local Similarity
                      81.0%;
 Matches 273; Conservative
                           31; Mismatches
                                           29;
                                               Indels
                                                           Gaps
                                                                  2;
          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
Qу
            1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
Db
         50 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
QУ
            61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQ1FNNILRRQIGTRTPTVEYI 120
Db
        120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
Qy.
               121 CTOONILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
Db
        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELII.DR 239
QУ
            181 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
Db
        240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQF 299
95
            141 HNFTIMTKYISKPENIKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQT 300
Db
         300 KLIEFLSSFOKERTDDEQFADEKNYLIKQIRDLKKTA 336
Qу
            301 KLIEFLSKFONDRTEDEOFNDEKTYLVKOIRDLKRAA 337
Db
RESULT 6
Q21543
                                    636 AA.
              PRELIMINARY;
                              PRT:
ID
    021643
AC.
    021643;
    01-NOV-1996 (TrEMBLrel. 01, Created)
DΤ
    01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT
    01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DT
    Hypothetical 72.3 kDa protein.
DE
    R02E12.2.
GN
    Caenorhabditis elegans.
OS
    Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC
    Rhabditidae; Peloderinae; Caenorhabditis.
OC
    NCBI TaxID=6239;
ÓΧ
RM
    [1]
    SEQUENCE FROM N.A.
R₽
RC
    STRAIN=Bristol N2;
    MEDLINE=99069613; PubMed=9851916;
RX
    None:
RA
    "Genome sequence of the nematode C. elegans: a platform for
RΤ
    investigating biology. The C. elegans Sequencing Consortium.";
RT
    Science 282:2012-2018(1998).
RL
    [2]
RN
RP
    SEQUENCE FROM N.A.
```

```
STRAIN=Bristol N2;
RC
    Leimbach D.;
RA
    "The sequence of C. elegans cosmid R02E12.";
RT
    Submitted (APR-1996) to the EMBL/GenBank/DDBJ databases.
RL
RN
    SEQUENCE FROM N.A.
RP
RC
    STRAIN=Bristol N2;
    Waterston R.;
RA
    "Direct Submission.";
RT
    Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
RL
DR
    EMBL: U53337; AAA96186.2; -.
    WormPep; R02E12.2; CE28410.
DR
    InterPro; IPR004892; Mo25.
DR
DR
    Pfam; PF03204; Mo25; 1.
KW
    Hypothetical protein.
             636 AA; 72282 MW; 85D5853E9E0E3193 CRC64;
SO
    SEQUENCE
                      62.6%; Score 1066.5; DB 5; Length 636;
 Query Match
                      60.4%; Pred. No. 6.3e-67;
 Best Local Similarity
 Matches 212; Conservative 53; Mismatches 69; Indels
                                                       17;
                                                           Gaps
                                                -QDKKTDKASEEVSKSLQ 47
          2 KKMP-LFSKSHKNPAEIVKILKDNLAILEK------
QУ
            258 KVMPLLFGKSHKSPADVVKTLREVLTILDKLPPPKLDKDGNIQSDKKYDKALDEVSKNVA 317
Db
         48 AMKEILCGTNEKEPPTE---AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNI 104 .
318 MIKSFIYGNDSAEPSSEHVVQVAQLAQEVYNANILPMLIKMLPKFEFECKKDVGQIFNNL 377
        183 LRRQIGTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKTILFSNQ 164
            338 LRRQIGTRSPTVEYLGARPEILIQLVQGYSVFDIALTCGLMLRESIRHDHLAKFILYSDV 437
100
        365 FRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVT 224
ŌΫ
                     438 FYTFFLYVQSEVFDISSDAFSTFKELTTRHKAIIAEFLDSNYDTFFAQYQNLLNSKNYVT 497
Db
         225 KRQSLKLLGELTLDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASP 284
Ç.7
            498 RRQSLKLLGELLLDRHNFNTMTKYISNPDNLRLMMELLRDKSRNIQYEAFHVFKVFVANP 557
Db
        235 HKTQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKT 335
Qу
            558 NKPKPISDILNRNREKLVEFLSEFHNDRTDDEQFNDEKAYLIKQIQEMKSS 608
Db
RESULT 7
08K312
                              PRT;
                                    205 AA.
              PRELIMINARY;
ID
    Q8K312
    Q8K312;
AC
    01-OCT-2002 (TrEMBLrel. 22, Created)
    01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT
    01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DT
    Similar to calcium binding protein, 39 kDa (Fragment).
DE
    Mus musculus (Mouse).
OS
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC
    Mammália; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC
```

```
NCBI TaxID=10090;
OX
RN
    [1]
ŔΡ
    SEQUENCE FROM N.A.
RA
    Strausberg R.;
    Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
RL
    EMBL; BC029053; AAH29053.1; -.
ĎΚ
    InterPro; IPR004892; Mo25.
DR
    Pfam; PF03204; Mo25; 1.
DR
                        1
    NON TER
                 1
FT
                                 015261A02F808169 CRC64;
              205 AA;
                      24582 MW;
    SEQUENCE
SO
                        51.3%; Score 875; DB 11;
                                                 Length 205;
  Query Match
  Best Local Similarity 83.6%; Pred. No. 4.8e-54;
                                                  Indels
                                                               Gaps -
  Matches 168; Conservative 17; Mismatches
                                              16:
         136 PQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHK 195
QУ
             1 PEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDIASDAFATFKDLLTRHK 60
Db
         196 VLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFAIMTKYISKPENL 255
QУ
             61 LLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDRHNFTIMTKYISKPENL 120
Dh
         256 KLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTDD 315
QΨ
             121 KLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQTKLIEFLSKFQNDRTED 180
Dh
         316 EQFADEKNYLIKQIRDLKKTA 336
             181 MOFNDEKTYLVKQIRDLKRAA 201
Db
RMSULT 0
Q3H5L9
                                       333 AA.
                                PRT:
     08Н5Ъ9
                PRELIMINARY;
ÍĐ
     O8H5L9;
AC
     01-MAR-2003 (TrEMBLrel. 23, Created)
DT
     01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT
     01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DT
     Putative MO25 protein (CGI-66).
DE
     OJ1060 D03.13.
GN
     Oryza sativa (japonica cultivar-group).
OS
     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC.
     Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
CC
     Ehrhartoideae; Oryzeae; Oryza.
OC
     NCBI_TaxID=39947;
OX
RN
     [1]
     SEQUENCE FROM N.A.
RР
     STRAIN=cv. Nipponbare;
RC
     Sasaki T., Matsumoto T., Yamamoto K.;
RΑ
     "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 7, BAC
RT
     clone:OJ1060 D03.";
RT
     Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
RL
     EMBL; AP003803; BAC22269.1; -.
DR
              333 AA; 38452 MW; CB6FC45E098C2401 CRC64;
     SEQUENCE
SQ
                        41.6%; Score 709.5; DB 10; Length 333;
```

Query Match

```
Best Local Similarity 44.0%; Pred. No. 3.7e-42;
 Matches 147; Conservative 67; Mismatches 109; Indels 11; Gaps
          6 LFSKSHKNPAEIVKILKDNLAILEKQ-----DKKTDKASEEVSKSLQAMKEILCGTNEK 59
Qу
                                        : ||::|: :: | |:
          4 LFKSKPRTPADVVRQTRELLIFLDLHSGSRGGDAKREEKMAELSKNIRELKSILYGNGES 63
Db
         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
ŨΛ
            64 EPVTEACVQLTQEFFRENTLRLLIICLPKLNLETRKDATQVVANLQRQQVSSKIVASEYL 123
ďG
         120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
             124 EANKDLLDTL1-SYENMDIALHYGSMLRECIRHQSIA-YVLESDHMKKFFDYIQLPNFDI 181
        180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYE-KLLQSENYVIKRQSLKLLGELILD 239
            182 ASDASATFKELLTRHKATVAEFLSKNYDWFFSEFNTRLLSSTNYITKRQAIKFLGDMLLD 241
Db
         239 RHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQ 298
Qy
             242 RSNSTVMMRYVSSKDNLMILMNLLRDSSKNIQIEAFHVFKLFAANKNKPTEVVNILVTNR 301
Db
         390 PKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDL 332
CY
             302 SKLLRFFAGFKTDK--DEQFEADKEQVIKEISAL 333
RESULT 9
@6L9L9
                                     345 AA.
               PRELIMINARY;
                               PRT:
    Q8L9L9
ΞĎ
AC
    Q8L9L9;
     01-OCT-2002 (TrEMBLrel. 22, Created)
    01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
TC
    01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
TC
    Hypothetical protein.
DE.
    Arabidopsis thaliana (Mouse-ear cress).
OS
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC
     Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC
     eurosids II; Brassicales; Brassicaceae; Arabidopsis.
CC
    MCBI_TaxID=3702;
XO
RM
\mathfrak{RP}
     SECUENCE FROM N.A.
     Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA
     Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RA
     "Full-length messenger RNA sequences greatly improve genome
RT
RT
     annotation.";
     Genome Biol. 0:0-0(2002).
RL
RN
     [2]
     SEQUENCE FROM N.A.
\mathbb{RP}
     Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA
RA
     Feldmann K.;
     "Pull-Length cDNA from Arabidopsis thaliana.";
ننظ
     Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
RL
     EMEL; AY088359; AAM65898.1; -.
DR
     InterPro; IPR004892; Mc25.
DR
     Pfam; PF03204; Mo25; 1.
DR.
```

```
Hypothetical protein.
ΚW
              345 AA; 39841 MW; 2C46A3D3DEBB47AA CRC64;
    SEQUENCE
SQ
                       39.4%; Score 671.5; DB 10; Length 345;
 Query Match
 Best Local Similarity 42.9%; Pred. No. 1.8e-39;
 Matches 140; Conservative 68; Mismatches 113; Indels
                                                                       2;
                                                               Gaps
          12 KNPAEIVKILKDNLAILEKQD----KKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVA 67
             QУ
          12 KTPQEVVKAIRDSLMALDTKTVVEVKALEKALEEVEKNFSSLRGILSGDGETEPNADQAV 71
Db
          68 QLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHPHILF 127
             \Omega \lambda
                                                        1:1
          72 QLALEFCKEDVVSLVIHKLHILGWETRKDLLHCWSILLKQKVGDTYCCVQYFEEHFELLD 131
Db
         138 MLLKGYEAPQTALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFATF 187. / W
                                    Qy
              1: 1: :||| || || ||||
         132 SLVVCYDNKEIALHCGSMLRECIKFPSLAKYILESACFELFFKFVELPNFDVASDAFSTF 191
Db
         188 KDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFAIMTK 247
             THILL TELL TO BE HELD THE HELD THE FOREST
Qy.
         192 KDLLTKHDSVVSEFLTSHYTEFFDVYERLLTSSNYVTRRQSLKLLSDFLLEPPNGHIMKK 251
Db
         248 YISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSS 307
             Çy
          %52 FILEVRYLKVIMTLLKDSSKNIQISAFHIFKIFVANPNKPQEVKIILARNHEKLLELLHD 311
58
          303 FOKER-TODEOFADEKNYLIKQIRDL 332
                 . .:|:|| :|| :|::|: |
          312 LSPGKGSEDDQFEEEKELIIEEIQKL 337
Db
 RESULT 10
 Q8LIF3
                                        322 AA.
                                 PRT;
                PRELIMINARY;
-TD
     Q8LTF3
 \mathbb{A}\mathbb{C}
     O8LIF3;
     01-OCT-2002 (TrEMBLrel. 22, Created)
 DΤ
     01-CCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT
     01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DT
     Hypothetical protein (P0503D09.26 protein).
 DE
     OJ1316_A04.9 OR P0503D09.26.
 GN
     Cryza sativa (japonica cultivar-group).
     Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Ċ$
     permatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 ОC
 OC
     Ehrhartoideae; Oryzeae; Oryza.
 OC
     NCBI_TaxID=39947;
 CX
      [1]
 RH
      SEQUENCE FROM N.A.
 RΡ
      STRAIN=cv. Nipponbare;
 RC
      Sasaki T., Matsumoto T., Yamamoto K.,
      "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, BAC
 RA
 RT
      clone:0J1316 A04.";
      Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
 RT
 RL
      [2]
 RN
      SEQUENCE FROM N.A.
 RΡ
      STRAIN=cv. Nipponbare;
 RC
      Sasaki T., Matsumoto T., Katayose Y.;
 RA
```

```
clone:P0503D09.";
RT
    Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
RL
    EMBL; AP003822; BAC06992.1; -.
DR
    EMBL; AP005455; BAC16736.1; -.
DR
    Gramene; Q8LIF3; -.
DR
    InterPro; IPR004892; Mo25.
DR
    Pfam; PF03204; Mo25; 2.
DR
    Hypothetical protein.
KW
    SEQUENCE 322 AA; 37091 MW; 99434DFA7C2DCD21 CRC64;
SO
                       34.6%; Score 590; DB 10; Length 322;
 Query Match
 Best Local Similarity 38.5%; Pred. No. 9e-34;
 Matches 129; Conservative 73; Mismatches 109;
                                                Indels
          4 MPLFSKSHKNPA----EIVKILKDNLAILEKQDKKTD-KASEEVSKSLQAMKEILCGTN 57
            1 MSFFFRAASRPARPSPQELVRSIKESLLAL---DTRTGAKALEDVEKNVSTLRQTLSGDG 57
Db
          58 EKEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVE 117
Ov
            |:
          58 EVEPNQEQVLQIALEICKEDVLSLFVQNMPSLGWEGRKDLAHCWSILLRQKVDEAYCCVQ 117
Db
         118 YISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTF 177
            118 YIENHFDLLDFLVVCYKNLEVALNCGNMLRECIKYPTLAKYILESSSFELFFQYVELSNF 177
         178 DIASDAFATFKOLLITRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELIL 237
            178 DIASDALNTFKDLLTKHEAAVSEFLCSHYEQFFELYTRLLTSTNYVTRRQSVKFLSEFLL 237 .
Dio :
         238 DRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKN 297
\Omega Y
                                               x : | || :|| : | :|: ||
         238 EAPMAQIMKRYIVEVSYLNIMIGLL-------KVFVANPNKPRDIIQVLVDN 282
         298 QPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDL 332
Qу
              :|::|: ::|| :|: :||:| |
         283 HRELLKLLGNLPTSKGEDEQLEEERDLIIKEIEKL 317
Db
RESULT 11
Q8K038
               PRELIMINARY;
                               PRT;
                                     103 AA.
ID
    Q8K038
АC
     01-OCT-2002 (TrEMBLrel. 22, Created)
DT
     01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT'
    01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DT
     Similar to RIKEN cDNA 1500031K13 gene.
DΕ
     Mus musculus (Mouse).
ОĊ
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC.
     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC
     NCBI TaxID=10090;
OX
R_{-1}
     [1]
     SEQUENCE FROM N.A.
RP
     TISSUE=Kidney;
RC:
     Strausberg R.;
RA
     Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
RL
```

"Oryza sativa nipponbare(GA3) genomic DNA, chromosome 7, PAC

RT

```
EMBL; BC034159; AAH34159.1; -.
DR
    InterPro; IPR004892; Mo25.
DR
    Pfam; PF03204; Mo25; 1.
DR
    SEQUENCE 103 AA; 11291 MW; EA86A9F6E9E426E0 CRC64;
SQ
                          25.5%; Score 435; DB 11; Length 103;
 Query Match
                         97.8%; Pred. No. 1.8e-23;
 Best Local Similarity
                                1; Mismatches 1;
                                                                 0;
                                                       Indels
                                                                     Gaps
          89; Conservative
            4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63
Q7
              1 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNDKEPPT 60
Db
           64 EAVAQLAQELYSSGLLVTLIADLQLIDFEGK 94
QУ
              61 FAVAQLAQELYSSGLLVTLIADLQLIDFEVK 91
RESULT 12
025188
                                          677 AA.
                                   PRT:
     025188
                 PRELIMINARY;
ID
     025188;
AC
     01-JAN-1998 (TrEMBLrel. 05, Created)
DT
     01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
     01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
     WMA topoisomerase I (TOPA). /
D\Xi
     BP0440.
GY.
     Melicobacter pylori (Campylobacter pylori).
00
     Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC
     Helicobacteraceae; Helicobacter.
610
\mathbb{K}^{0}
     NCBI_TaxID=210;
     [1]
\mathbb{P}_{+}
     SEQUENCE FROM N.A.
23.2
     STRAIN=26695 / ATCC 700392;
РC
     MEDLINE=97394467; PubMed=9252185;
\overline{A}X
     Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
\Sigma A
     Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
AS
     Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA
     Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA
     McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RА
     Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA
     Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA
     Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA
RA
     "The complete genome sequence of the gastric pathogen Helicobacter
\mathbb{R}T
     pylori.";
\mathbf{RT}
     Nature 388:539-547(1997).
RL
     EMBL; AE000559; AAD07502.1; -.
DR
     TIGR; HP0440; -.
DR
     InterPro; IPR003601; DNAtopI_ATP bind.
DR
     InterPro; IPR003602; DNAtopI_DNA bind.
DŘ
     InterPro; IPR000380; DNA_tpisomrase.
DR
     InterPro; IPR006171; Toprim_dom.
DR
     InterPro; IPR006154; Toprim_sub.
DR_
     Pfam; PF01131; Topoisom_bac; 1.
_{
m DR}
     Pfam; PF01751; Toprim; 1.
DR
     PRINTS; PRO0417; PRTPISMRASEI.
DR
```

```
SMART; SM00436; TOP1Bc; 1.
DR
    SMAR'I; SM00493; TOPRIM; 1.
DR
    Hypothetical protein; Isomerase; Complete proteome.
KW
    SEQUENCE 677 AA; 77677 MW; 4B285B81F1092BB4 CRC64;
SO
                     7.9%; Score 134.5; DB 16; Length 677;
 Ouery Match
 Best Local Similarity 21.6%; Pred. No. 0.24;
 Matches 88; Conservative 58; Mismatches 134; Indels 127; Gaps
          7 FSKSHKNPA-EIVKILKDNL------AILEKQDKK---TDKASEEVSKSLQAMKE 51
QУ
               222 FKFKDKNEASQFLKDLKDGLGSMSVLVSLKESLSNKEPKKPFTTSKLLSQASKSLKI--- 278
         52 ILCGTNEKEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGT 111
QУ
                   -----PTKEIAQLAQKLFEAGLITYHRTDSEFLSPEYLKEHEVFFEPIY----- 322
Эþ
        112 RSPTV----EYIS------AHPHILFMLLKGYEAPQIALRCGIMLRECTRHE 153
QУ
                                  ]:| || :
        323 -- PSVYQYREYKAGKNSQAEAHEAIRITHPHALKDLEKVCSDAKISEELALKLYQLIYTN 380
Db
        184 PL---AKIILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIF 210-
QУ
            ORG TICSOSRNALY-NOYDCIFK-----IKSESFKLSFKLKEKGFLEIEELIOGKEEIN 43153
        EDYEKTLQSENYVTKRQSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQ 270; 1
            432 RE-EQESEIENFSLKENDSVPLKEVFIKK-----IEKPSPKPYKESAFIPLLESEG---- 481
        PVE FEAFHVFKVFVASPHKTQPIVEILLKNQ------PKLIEFLSSFQKERTDD- 315
                : | : | | : | | : |
        432 -----IGRPSTYASFLDLLLKRKYISIDTKTNAITPTSQGLEVISFFKKDKEVDF 531
         316 ----EQF------ADEKNYLIKQIRDLKKTA 336
                        532 IALTSKDKSKLGNTTKQFEECLDLIMRGEASYEKFMLEVISKLKSTA 578
RESULT 13
026049
                             PRT;
                                   430 AA.
              PRELIMINARY;
    026049
ΞĎ
AC
    026049;
    01-JAN-1998 (TrEMBLrel. 05, Created)
DT
    01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT
    01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DT
    Hypothetical protein HP1520.
DE
    HP1520.
GN
    Helicobacter pylori (Campylobacter pylori).
CS
    Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacterales;
OC
    Helicobacteraceae; Helicobacter.
OC.
    NCBI TaxID=210;
ОХ
RN
     [1]
     SEQUENCE FROM N.A.
RP
     STRAIN=26695 / ATCC 700392;
RC
     MEDLINE=97394467; PubMed=9252185;
РX
     Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
 RA
```

SMART; SM00437; TOP1Ac; 1.

DR

```
Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
    Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA
    Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA
    McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA
    Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA
    Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA
    Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
ŘΑ
    Venter J.C.;
RA
    "The complete genome sequence of the gastric pathogen Helicobacter
RT
RT
    Nature 388:539-547(1997).
RL
    EMBL; AE000650; AAD08565.1; -.
DR
    TIGR; HP1520; -.
DR
    Hypothetical protein; Complete proteome.
KW
            430 AA; 50573 MW; 23DC6FE5E956B629 CRC64;
SO
                       7.5%; Score 128; DB 16; Length 430;
 Query Match
                      20.9%; Pred. No. 0.39;
 Best Local Similarity
 Matches 82; Conservative 73; Mismatches 135; Indels 102; Gaps
                                                                  20;
          7 FSKSHKNPAEI----VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPP 62
Qy:
            60 FYPNRKSKIEIEFNGEKILKENVAVFHSYDE--EFSSEDSVTTFMAKSDL-----KQQY 111
         53 TEAVAQLAQELYSSGLLVTL--IA----- DLQLIDFEGKKDVTQIFNNILR----- 106
ŨУ
                      112 DNILLELEKE--KKALLKSLRDIASGFDYEEEIKTIKNEKNKSFYEILDNHLTEIESSEK 169
ot.
         107 -----RQIGTRSPTV-EYISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKII 159
                                                          | | ::
                 --ELLSQSK 211
         10 HYSFKYRDIFDGSKKVKDFVNKHHDLIEQYFNKYQ---
Dis
         1.50 LF-----SNQFRDFFKYVELSTFDTASDAFATFKDLLTRHKVLVADFLEQ-----
            212 IFKHMNSGDFGTNHADDLKKALENNRFFKANHSLKIAGEEITNYQKL-SDIFENEKNRIL 270
L).5
         205 NYDTIFEDYEKLLQSENYVTKRQSLKLLGELI------LDRHNF--AIMTKYISKP 252
            271 NNEELKESFDKI---EKVINANKELKAFKDAISKDNTLLTEFLDYDSFRKKVLFSYLKQV 327
Db
         253 -ENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKE 311
QУ
             328 IQNVKSLVNLYREKKPEIE----EIIKQASKDQKEWESVIEIF--NQRFLVPFKVELQNQ 381
Db
         312 R----TDDEQ----FADEKNYLIKQIRDLKK 334
          382 KDILLNKDAAQFRFIFSDDNQDMNVQKEDLQK 413
RESULT 14
EUXWQQ
                               PRT;
               PRELIMINARY;
ID
     09WXU3
20
     09WXU3;
     01-NOV-1999 (TrEMBLrel. 12, Created)
DΤ
     01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT
     01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DT
     COME protein, putative.
DE
```

RA

```
GN
    TM0088.
    Thermotoga maritima.
0S
    Bacteria; Thermotogae; Thermotogales; Thermotogaceae; Thermotoga.
OC
    NCBI TaxID=2336;
OX
    [1]
RN
    SEQUENCE FROM N.A.
RP
    STRAIN=MSB8 / DSM 3109;
RC
    MEDLINE=99237316; PubMed=10360571;
RX
    Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA
    Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA
    McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA
    Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
    Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA
RA
    Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RA
    "Evidence for lateral gene transfer between Archaea and Bacteria from
RT
    genome sequence of Thermotoga maritima.";
RT
    Nature 399:323-329(1999).
RL
    EMBL; AE001695; AAD35182.1; -.
DR
    TIGR; TM0088; -.
DR.
    InterPro; IPR004846; GSPII/IIIprotein.
DR
    InterPro; IPR001993; Mitoch_carrier.
DR
    Pfam; PF00263; GSPII III; 1.
DR
    PROSITE; PS00215; MITOCH CARRIER; 1.
DR
     Complete proteome.
ΚW
    PROUENCE 1285 AA; 145209 MW; 057435F821FB0EA5 CRC64;
SQ
                        7.2%; Score 123.5; DB 16; Length 1285; 2000
  One my Matith
  Best Local Similarity 21.5%; Pred. No. 3;
                                                            (= 1 · - =)
  Matches 86; Conservative 78; Mismatches 129; Indels 107; Gaps 23;
           1 MKKMPLFSKSHKNPAEIVKILKDNLAILEKQD-----KKT----DKASEEV---SKS 45
            556 LKVAMLSGKEEEN-----VQKAAEELQIISSEERIIRFVKKTENVPIDKAKNVVLQLYSVS 611
                                                        --LVTLIAD--
          46 LQAMKEILCGTNEKEPPTEAVAQLAQELYSSGL-----
             612 IEELGNELVVIGERE-EVEKAADLLQKIFSSEVEISRDFVKLPSWIDEQEKLLEVVKNSA 670
 Dρ
          36 ---LQLID----FEGKXD----VTQIFNNILRRQIG--TRSPTVEYI---SAHPHILFML 1.29
                     Q_{\mathcal{I}}
         671 GITYEILDGVVYFEGTKENVEKAKELFSDIVEK-LGEVRKEETVEFLEVNSSFPVDEFIN 729
 Db
         330 LKGYEAPQIALRCGIMLRECIRHEPLAKIIL-----FSNQFRDFF----KYVELST 176
                                                : | ::|
             1 | | :
         730 LSGKLYPDVT------CFSLDQLGLLVLKGSSEAVEDLSSMYRSFFERHQKIVKENV 780
 Do
         177 FD---IASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLG 233
           OY
          781 FDRLMLEVPSGFSFEEFKTFLEVLVPEVKQ----VVYLDKLNLLLVEVPVSQSERVKSLL 836
 Db
          234 ELILDRHNFAIMTKYIS------KPENL-KLMMNLLRDKSPNIQFEAF-HVFKVFVAS 283
 QΥ
             : | : | : |
                                   837 DTFLKKEEAVSEKKAVKSVTIPSGVNPDELSSYLKKLLR----NVEITVFPNMGQMIVEG 892
 d\mathbf{Q}
          284 P-HKTQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEK 322
 QΫ
           393 PENEVEKAVELVEAEKEKIV-----LKERKDYVKVSDGK 926
 Dh
```

```
RESULT 15
Q58914
                                         1175 AA.
                                   PRT;
                 PRELIMINARY;
     Q58914
ŢD
     Q58914;
AC
     01-JUN-1998 (TrEMBLrel. 06, Created)
DT
     01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT
     01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DT
     Hypothetical protein MJ1519.
DΕ
GN
     MJ1519.
     Methanococcus jannaschii.
OS
     Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC
     Methanocaldococcaceae; Methanocaldococcus.
OC.
     NCBI TaxID=2190;
OX
RN
     SEQUENCE FROM N.A.
RP
     STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RC
     MEDLINE=96337999; PubMed=8688087;
RX
     Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA
     Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RΑ
     Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
     Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA
RΑ
     Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA
     Otterback T.P., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
7.3.
     Cotter M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
\mathbb{R}\mathbb{A}
     Mlenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
22.
     "Complete genome sequence of the methanogenic archaeon, Methanococcus
 20.0
      jannəschii.";
. . . . . .
      Science 273:1058-1073(1996).
ŔĽ
     MBL; U67593; AAB99538.1; -.
DR
     TIGR; MJ1519; -.
DP.
      InterPro; IPR003593; AAA ATPase.
 J.R.
      SMART; SM00382; AAA; 1.
\mathbb{DR}
      Hypothetical protein; Complete proteome.
 ANN.
                 1175 AA; 138618 MW; 99082EA5A4D11140 CRC64;
 ŽΟ
      SEQUENCE
                                   Score 120; DB 17; Length 1175;
                            7.0%;
   Query Mauch
                           21.5%; Pred. No. 4.8;
   Best Local Similarity
                                                                        Gaps 15;
             76; Conservative 58; Mismatches 131;
                                                         Indels
                                                                   88;
             7 FSKSHKNPAEIVKILKD-NLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEA 65
 QV
                                           | | | : | : | : ::
                                                                 | | | |
                   232 FNKFREENQDFDKYLTDENIAFRPHVMKKFDEFAENIKKVIAELE----GSKYKYPGLPG 287
 Db
            66 VAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHPHI 125
 ÇУ
                                                          : |
                                                                   : :
                                                :::
                                          11:1
                                    ::[
           288 V-----LYFLGMEDAYSRYIELWKNEGEKGEEKLYNALI-ESLENRKENLEF----- 333
 a'C
           126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFK----YVELSTFDIA- 180
 ÇΊ
                                                       :||:||
                                                                  | | | | :
                                      -GITKKVIDKFIAQKEEFREFLKNYAVYYELSAFKLEK 370
 Dic
           181 ----SDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSL----- 229
 Qy
                                                  : :|::: |
                                    : | | | | : : |
                             :::
           371 IKEQYEKEFINLDNIIKNPYILVED-IKEN-----DSFERIIFEELDSWERRRLGDKFNP 424
 Db
```

Ov	230KLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAF 274
Qу	
Db	425 YSPYRVRALLVE-ILKRHLSSGNTTISTKDLKDFFEKMDKDIVKITFDEFLRII 477
Ov	275 HVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIK 327
	::   : : :  :         : ::   :
Db	478 EEYKDIISEKVEIVKKEVKNNENKEIIELFTLKEIREYEEIIENTINYLLK 528

Search completed: January 7, 2004, 16:48:05 Job time : 56 secs

# GenCore version 5.1.6 Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

January 7, 2004, 16:44:17; Search time 17 Seconds

(without alignments)

932.235 Million cell updates/sec

Title:

US-10-088-872-2

Perfect score:

1704

Sequence:

1 MKKMPLFSKSHKNPAEIVKI......FADEKNYLIKQIRDLKKTAP 337

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched:

127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters:

127863

Minimum DE seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0% Maximum Match 100%

Listing first 45 summaries

Database :

SwissProt\_41:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, we gr and is derived by analysis of the total score distribution.

## SUMMARIES

		ક				
Result		Query			× D	Description
Mo.	Score	Match	Length	DB	ID	
	± 1£685	98.9	334	1	MO2L_HUMAN	Q9h9s4 homo sapien:
2	±565 ±569		334	1	MO2L MOUSE	Q9db16 mus musculu
2	1381	81.0	341	1	MO25_HUMAN	Q9y376 homo sapien
	1376	80.8	341	1	MO25_MOUSE	Q06138 mus musculu
14 5	1111	65.2	339	1	MO25 DROME	P91891 drosophila
	1006.5	59.1	338	1	MO2M_CAEEL	018211 caenorhabdi
7	\$34.5	49.0	329	1	YFV6 SCHPO	Q9p7q8 schizosacch
<i>,</i> ទំ	776	45.5	321	1	DE76_CHLPR	Q9xfy6 chlorella p
9	728	42.7	343	1	MO2N_ARATH	Q9fgk3 arabidopsis
10	716.5	42.0		1	MO2M ARATH	Q9m0m4 arabidopsis
11	665	39.1			HYMA EMENI	060032 emericella
12	53 <b>2</b>	37.1			MO2L ARATH	Q9zq77 arabidopsis
13	485	28.5	_		HYM1_YEAST	P32464 saccharomyc
14	143.5	8.4			MO2L CAEEL	Q9tzm2 caenorhabdi:
15	128.5	7.5			AKA9 HUMAN	Q99996 h a-kinase
16	125.5	7.4			Y295 RICPR	Q9zdn2 rickettsia
17	118.5	7.0			DPO5_SCHPO	060094 schizosacch

13	116.5	6.8	724	1	HMMR HUMAN	075330	homo sapien
19	115	6.7	474	1	GSHB_MOUSE	P51855	mus musculu
20	112.5	6.6	1411	1	YM42_YEAST	Q03214	saccharomyc
21	109.5	6.4	978	1	RA50 AQUAE	067124	aquifex aeo
22	109	6.4	695	1	YCX7 CHLVU	020159	chlorella v
23	109	6.4	1401	1	LATA LATMA	P23631	latrodectus
24	108.5	6.4	586	1	2A5D_RABIT	Q28653	o serine/th
25	108.5	6.4	602	1	2A5D HUMAN	Q14738	h serine/th
26	108.5	6.4	1075	1	Y124 METJA	Q57588	methanococc
27	108	6.3	568	1	DNAB PORPU	P51333	porphyra pu
28	107.5	6.3	483	1	ACPA BACAN	Q44643	bacillus an
29	107.5	6.3	1042	1	T1RH_METJA	Q60295	methanococc
30	1.07.5	6.3	1726	1	MSP1 PLAFC	P04934	plasmodium
31	107.5	6.3	1726	1	MSP1 PLAFP	P50495	plasmodium
32	107	6.3	1727	1	ALM1_SCHPO	Q9utk5	schizosacch
33	106	6.2	474	1	GSHB_HUMAN	P48637	homo sapien
34	105.5	6.2	793	1	REGA_DICDI	Q23917	dictyosteli
35	105.5	6.2	847	1	RSG2_RAT		rattus norv
3,6	104.5	6.1	1701	1	MSP1_PLAFF	P13819	plasmodium
37	104.5	5.1	1701	1	MSP1_PLAFM		plasmodium
38	104	6.1	859	1.	MUTS_AQUAE		aquifex aeo
39	104	6.1	1290	1	RA50_SCHPO	Q9utj8	schizosacch
4.0	104	5.1	1682	1	MSP1_PLAF3	P19598	plasmodium
41	103.5	6.1	641	.1	PRIM_UREPA		ureaplasma
42	103	5.0	2663	1	CENE_HUMAN		homo sapien
43	102.5	6.0	502	1.	URIC_BACSB	Q45697	bacillus sp
44	102.5	5.0	975	1	KINH_DROME	917210	drosophila
45	102.5	5.0	1:202	1	RFM2_YEAST	Q02773	saccharomyc

## ALIGNMENTS

```
MO2L HUMAN
                                    PRT;
                                           334 AA.
                    STANDARD;
     MO2L HUMAN
     Q9H9S4; Q9BZ33;
AC
     16-OCT-2001 (Rel. 40, Created)
DT
     16-OCT-2001 (Rel. 40, Last sequence update)
ÐΤ
     28-FEB-2003 (Rel. 41, Last annotation update)
DT.
DΕ
     MO25-like protein.
     Homo sapiens (Human).
OS
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC.
     Mammalia; Eutheria: Primates; Catarrhini; Hominidae; Homo.
OC
OX
     NCBI TaxID=9606;
RN
     [1]
RΡ
     SEQUENCE OF 4-334 FROM N.A.
     Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA
     Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
R.A
     Waqatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
F_{\alpha}
     Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA
     Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
AS
     Yamanoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA.
     Ninomiya K., Iwayanagi T.;
RA
     "NEDO human cDNA sequencing project.";
RT
     Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
RL
RN
     [2]
```

RESULT 1

```
SEQUENCE OF 276-334 FROM N.A.
RP
    Pearce A.;
RA
    Submitted (JAN-2001) to the EMBL/GenBank/DDBJ databases.
RL
    -!- SIMILARITY: Belongs to the Mo25 family.
CC
    _____
CC
    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC
    the European Bioinformatics Institute. There are no restrictions on its
CC
    use by non-profit institutions as long as its content is in no way
CC
    modified and this statement is not removed. Usage by and for commercial
CC
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC
    or send an email to license@isb-sib.ch).
CC
CC
    EMBL; AK022639; BAB14147.1; ALT_INIT.
DR
    EMBL; AL138875; CAC28084.1; -.
DR
    InterPro; IPR004892; Mo25.
DR
    Pfam; PF03204; Mo25; 1.
DR
            334 AA; 38728 MW;
                             97702273D8548432 CRC64;
    SEQUENCE
SQ
                      98.9%; Score 1685; DB 1; Length 334;
  Query Match
                      99.7%; Pred. No. 1.3e-100;
  Best Local Similarity
                                               Indels
                                                        0;
                           0; Mismatches
                                           1:
  Matches 333; Conservative
          4 -MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63
Qу
            1 MPLPSKSHKNPAEIVKELKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 60
Db
         64 EAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHP 123
QУ
            51 EAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHP 120
Db
         124 HILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDA 183
QY
            121 HILFMLLKGYEAPQIALRCGIMLRECIRHEPLVKIILFSNQFRDFFKYVELSTFDIASDA 180
Db
         134 FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFA 243
ÒΆ
            181 FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFA 240
Db
         244 IMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIE 303
QУ
            241 IMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIE 300
DD
         304 FLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
Qy
            301 FLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 334
Db
RESULT 2
MO2L MOUSE
                                    334 AA.
                              PRT:
                 STANDARD;
     MO2L MCUSE
ΞD
     Q9DB16; Q8BG52; Q91WB8; Q91YL0;
AC
     16-CCT-2001 (Rel. 40, Created)
DT
     15-SEP-2003 (Rel. 42, Last sequence update)
     15-822-2003 (Rel. 42, Last annotation update)
DT
     MO25-like protein.
DΕ
     Mus musculus (Mouse).
OS
```

```
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC
     Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC
OX
     NCBI TaxID=10090;
RN
     [1]
     SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RР
RC
     STRAIN=C57BL/6J;
RC
     TISSUE=Cerebellum, Eye, Pituitary, and Testis;
     MEDLINE=22354683; PubMed=12466851;
RX
     Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RΑ
     Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA
     Yaqi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA
     Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
ŖΑ
     Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA
     Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA
     Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA
     Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA
     Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA
     Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA
     Konaqaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA
     Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RR
     Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA
     Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
ŖΑ
     Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA
     Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K., 🕟
ESI
     Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
PA
     Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.
\mathbb{R} A
     Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
\mathbb{R}\mathbf{A}
     Yuan Z.. Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
PA
     Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA
     Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
\mathbb{R}\mathbb{A}
     Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA
     Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
ÎA
     Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA
     Birney E., Hayashizaki Y.;
RA
RT
     "Analysis of the mouse transcriptome based on functional annotation of
RT
     60,770 full-length cDNAs.";
     Nature 420:563-573(2002).
RL
RN
     SEQUENCE FROM N.A. (ISOFORM 1).
RΡ
     STRAIN=FVB/N; TISSUE=Mammary gland, and Salivary gland;
RC
     MEDLINE=22388257; PubMed=12477932;
RX
     Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA
     Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA
     Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA
     Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
PA
     Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RЯ.
     Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
2.7
     Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA
     Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA
     Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
R\Delta
     Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
R^{A}
     Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., 🕐
RA
     Pahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA
     Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA
     Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
F.E.
     Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA
     Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
```

RA

```
"Generation and initial analysis of more than 15,000 full-length human.
RT
    and mouse cDNA sequences.";
RT
    Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RL
    -!- ALTERNATIVE PRODUCTS:
CC
       Event=Alternative splicing; Named isoforms=2;
CC
CC
         IsoId=Q9DB16-1; Sequence=Displayed;
CC
CC
       Name=2:
         IsoId=Q9DB16-2; Sequence=VSP_007417, VSP_007418;
CC
         Note=No experimental confirmation available;
CC
    -!- SIMILARITY: Belongs to the Mo25 family.
CC
    _______
CC
    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC
    the European Bioinformatics Institute. There are no restrictions on its
CC
    use by non-profit institutions as long as its content is in no way
CC
    modified and this statement is not removed. Usage by and for commercial
CC
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC
    or send an email to license@isb-sib.ch).
CC
   CC
    EMBL; AK005323; BAB23953.2; ALT_INIT.
DR
    EMBL; AK030474; BAC26978.1; ALT_INIT.
DR
    MMBL; AK053642; BAC35457.1; ALT_INIT.
    EMBL: AK076758; BAC36470.1; ALT INIT.
DR
    DR
    EMBL; BCC16128; AAH16128.1; -.
DR
    EMBL; BC016546; AAH16546.1; -. >
DR
   MGD; MGI:1916258; 1500031K13Rik.
\mathbb{T}\mathbb{R}
    InterPro; IPR004892; Mo25.
\mathbb{D}\mathbb{R}
   Pfam; PF03204; Mo25; 1.
DR
    Alternative splicing.
XH
                              VFVASPHKTQPIVEILLK -> NSVFITNRIHGLKRWDSS
   VARSPLIC 276
                     293
\mathbb{P}^{n}\Gamma
                              (in isoform 2).
VT.
                              /FTId=VSP 007417.
FT
                              Missing (in isoform 2).
    VARSPLIC 294
                     334
\mathbb{F}^{\mathrm{d}}\Gamma
                              /FTId=VSP 007418.
FT
                              S -> P (IN REF. 1; BAB23953).
PT
    CONFLICT
               42
                     42
                         L -> R (IN REF. 2; AAH16546).
               229
                     229
PT
    CONFLICT
    SEQUENCE 334 AA; 38718 MW; 822F04A87FB4EB6F CRC64;
SQ/
                       97.9%; Score 1669; DB 1; Length 334;
  Query Match
 Best Local Similarity 98.5%; Pred. No. 1.4e-99;
  Matches 329; Conservative 2; Mismatches 3; Indels
           4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63
            1 MPLFSKSHKNPAEIVKTLKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNDKEPPT 60
D.b
          E4 EAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHP 123.
             51 EAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRCPTVEYISSHP 120
         124 HILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDA 183
Qу
             121 HILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDA 180
Db
```

Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

```
184 FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFA 243
              QУ
          181 FATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFT 240
Db
          244 IMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIE 303
              Qγ
          241 IMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIE 300
Db
          304 FLSSFQKERTDDEQFADEKNYLIKQIRDLKKTAP 337
QУ
              301 FLSSFQKERTDDEQFADEKNYLIKQIRDLKKAAP 334
Db
RESULT 3
MO25 HUMAN
                                   PRT;
                    STANDARD;
     MO25 HUMAN
TD
     09Y376;
AC
     16-OCT-2001 (Rel. 40, Created)
DT
     16-OCT-2001 (Rel. 40, Last sequence update)
DT
     15-SEP-2003 (Rel. 42, Last annotation update)
D_{c\bar{c}}
     MO25 protein (CGI-66).
DΕ
     MO25.
GN
     Homo sapiens (Human).
     Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OS
OC
     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC
     NCBI TaxID=9606;
 OX
      [2]
 RM
      SEQUENCE FROM N.A.
 RP
     MEDILINE=20272150; PubMed=10810093;
 \mathbb{P}\mathbb{X}
     Lai C.-H., Chou C.-Y., Ch'ang L.-Y., Liu C.-S., Lin W.-C.;
 7. A.
      Adentification of novel human genes evolutionarily conserved in
 \mathbb{R}^{r_0}
      Caenorhabditis elegans by comparative proteomics.";
 \mathbb{R}^{n}
      Genome Res. 10:703-713(2000).
 13.1
 MSI
      SEQUENCE FROM N.A.
 \mathbf{PP}
      TISSUE=Hypothalamus;
 RC
      Jin W., Shi J., Ren S., Gu J., Fu S., Huang Q., Dong H., Yu Y., Fu G.,
 RA
      Wang Y., Chen Z., Han Z.;
 RA
      "A novel gene expressed in the human hypothalamus.";
 RT
      Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.
 RL
 RM
      [3]
      SEQUENCE FROM N.A.
 D.O
      TISSUE=Duodenum;
 æc
      MEDLINE=22388257; PubMed=12477932;
 \mathbb{R}X
      Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
      Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 \mathbb{R}\mathbb{Z}.
      Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 \mathbb{R}\mathsf{A}
      Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 PA
 RA
      Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
      Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA
      Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
  RA
      Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
      Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
  RA
      Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
  RS
      Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
  RA
       Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
  RA
       Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
  RA
  RA
```

```
Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA
   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA
    Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA
    Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
    "Generation and initial analysis of more than 15,000 full-length
RA
RT
    human and mouse cDNA sequences.";
RT
    Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RL
    -!- SIMILARITY: Belongs to the Mo25 family.
    ______
CC
    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC
    the European Bioinformatics Institute. There are no restrictions on its
CC
    use by non-profit institutions as long as its content is in no way
CC
    modified and this statement is not removed. Usage by and for commercial
CC
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC
CC
    or send an email to license@isb-sib.ch).
CC
CC
    EMBL; AF151824; AAD34061.1; -.
DR
    EMBL; AF113536; AAF14873.1; -.
DR
    EMBL; BC020570; AAH20570.1; -.
DR.
    InterPro; IPR004892; Mo25.
DR
    Pfam; PF03204; Mo25; 1.
DR
    SEQUENCE 341 AA; 39869 MW; EC710A528B6F9811 CRC64;
                      81.0%; Score 1381; DB 1; Length 341;
  Query Match
  Best Local Similarity 81.0%; Pred. No. 3.1e-81;
 Matches 273; Conservative 31; Mismatches 29; Indels
                                                          Gaps -
          4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
            1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
Db
         60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
            ΟУ
         61 EPQTEAVAQLAQELYNSGLLSTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
OO
         120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
                Qу
         121 CTQQNILFMLLKGYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
Tio.
         180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239.
            Oy.
         131 ASDAFATFKDLLTRHKLLSAEFLEQHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
 D
         240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
            Qy
         241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQA 300.
 Œ
         300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTA 336
 Qν
             301 KLIEFLSKFQNCRTEDEQFNDEKTYLVKQIRDLKRPA 337
 Db
```

RESULT 4
M025\_MOUSE
ID M025\_MOUSE STANDARD; PRT;
AC Q06138;

```
01-FEB-1994 (Rel. 28, Created)
DT
    01-FEB-1994 (Rel. 28, Last sequence update)
DT
    16-OCT-2001 (Rel. 40, Last annotation update)
DΤ
    MO25 protein.
    MO25 OR CAB39.
GN
    Mus musculus (Mouse).
OS
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC
    NCBI TaxID=10090;
OX
    [1]
RN
    SEQUENCE FROM N.A.
RΡ
    MEDLINE=93119656; PubMed=8418809;
RX
    Mivamoto H., Matsushiro A., Nozaki M.;
RA
    "Molecular cloning of a novel mRNA sequence expressed in cleavage
RT
    stage mouse embryos.";
RT
    Mol. Reprod. Dev. 34:1-7(1993).
RL
    -!- FUNCTION: ONE OF THE FIRST GENES TO BE TRANSCRIBED DURING MOUSE
CC
        DEVELOPMENT, MAY PLAY SOME GENERAL FUNCTION.
CC
    -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC
    -!- DEVELOPMENTAL STAGE: TRANSCRIBED DURING EARLY MOUSE DEVELOPMENT.
CC
        DETECTED AT ALL DEVELOPMENTAL STAGES FROM THE EGG THROUGH THE
CC
        BLASTOCYT, MOST ABUNDANT AT THE 2-CELL STAGE.
CC
    -!- SIMILARITY: Belongs to the Mo25 family.
CC
     ______
CC
    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC
    the European Bioinformatics Institute. There are no restrictions on its
CC
    use by non-profit institutions as long as its content is in no way
CC
    modified and this statement is not removed. Usage by and for commercial
CC
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC
    or send an email to license@isb-sib.ch).
CC
     CC
    EMPL; S51858; AAB24301.1; -.
DR
    MGD: MGI:107438; Cab39.
DR
    InterPro; IPR004892; Mo25.
DR
    Pfam; PF03204; Mo25; 1.
DR^{\circ}
    SEQUENCE 341 AA; 39842 MW; E7F668529D6FE811 CRC64;
SO.
  Query Match 80.8%; Score 1376; DB 1; Length 341; Best Local Similarity 80.7%; Pred. No. 6.5e-81;
  Matches 272; Conservative 32; Mismatches 29; Indels
                                                          4; Gaps
           4 MPL-FSKSHKNPAEIVKILKDNLAILEKQ---DKKTDKASEEVSKSLQAMKEILCGTNEK 59
Öλ
             1 MPFPFGKSHKSPADIVKNLKESMAVLEKQDISDKKAEKATEEVSKNLVAMKEILYGTNEK 60
Db
          50 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
Qy.
             61 EPQTEAVAQLAQELYNSGLLGTLVADLQLIDFEGKKDVAQIFNNILRRQIGTRTPTVEYI 120
Db
         120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
ĞΣ.
                121 CTQQNILFMLLKCYESPEIALNCGIMLRECIRHEPLAKIILWSEQFYDFFRYVEMSTFDI 180
Db
         180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
Ωy
             181 ASDAFATFKOLLTRHKLLSAEFLECHYDRFFSEYEKLLHSENYVTKRQSLKLLGELLLDR 240
Db.
```

```
240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
Qу
              241 HNFTIMTKYISKPENLKLMMNLLRDKSRNIQFEAFHVFKVFVANPNKTQPILDILLKNQT 300
Db
          300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKKTA 336
ÔУ
            301 KLIEFLSKFQNDRTEDEQFNDEKTYLVKQIRNLKRAA 337
Db
RESULT 5
MO25 DROME
                                          339 AA.
                                   PRT;
                    STANDARD;
     MO25 DROME
     P91891; Q9VV85;
AC
     16-OCT-2001 (Rel. 40, Created)
DT
     16-OCT-2001 (Rel. 40, Last sequence update)
DT
     16-OCT-2001 (Rel. 40, Last annotation update)
ÐТ
     MO25 protein (dMo25).
DE
     MO25 OR CG4083.
CN
     Drosophila melanogaster (Fruit fly).
OS.
     Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC'
     Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC
     Ephydroidea; Drosophilidae; Drosophila.
OC
OX
     NCBI TaxID=7227;
RN
     [1]
     SEQUENCE FROM N.A.
\mathbb{R}P
     MEDLINE=96268479; PubMed=8672247;
RE
     Nozaki M., Onishi Y., Togashi S., Miyamoto H.;
\mathbb{A}\mathbb{C}
     "Molecular characterization of the Drosophila Mo25 gene,
                                                              which is
RT
     conserved among Drosophila, mouse, and yeast.";
RT
     DNA Cell Biol. 15:505-509(1996).
TYTE.
\mathbb{R}\mathbb{N}
     [2]
     SEQUENCE FROM N.A.
2P
     STRAIN=Berkeley;
RC
     MEDLINE=20196006; PubMed=10731132;
RX
     Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA
     Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RΑ
     George R.A., Lewis S.E. Richards S., Ashburner M., Henderson S.N.,
RA
     Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA
     Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA
     Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA
     Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA
     Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
PA
     Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA
     Porkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RΑ
     Eurtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA
     Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA
     de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA
      Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA
     Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA
      Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA
      Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA
      Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA
     Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA
      Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA
      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RΑ
      Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA
```

```
Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA
    Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA
    Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA
    Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA
    Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA
    Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA
    Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA
    Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA
    Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA
    Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA
    Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RΑ
    Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA
    Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA
    Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA
    "The genome sequence of Drosophila melanogaster.";
RT'
    Science 287:2185-2195(2000).
RL
    -!- SIMILARITY: Belongs to the Mo25 family.
CC
    _____
CC
    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC
    the European Bioinformatics Institute. There are no restrictions on its
CC
    use by non-profit institutions as long as its content is in no way
CC
    modified and this statement is not removed. Usage by and for commercial
CC
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC.
    or send an email to license@isb-sib.ch).
CC
CC
    EMBL; AB000402; BAA19098.1; ---
DR
    EMBL; AE003526; AAF49432.1; -.
ÐΒ
    FlyBase; FBgn0017572; Mo25.
DR-
    InterPro; IPR004892; Mo25.
DR
                                                           . .
    Mfam; PF03204; Mo25; 1.
(i,j)
                              Y \rightarrow H (IN REF. 1).
    CONFLICT 51 51
315
                           V -> L (IN REF. 1).
    CONFLICT
               102
                     1.02
ery)
              339 AA; 39385 MW; 5790BD91754C1C74 CRC64;
     SEQUENCE
30
                       65.2%; Score 1111; DB 1; Length 339;
  Query Match
                       65.0%; Pred. No. 4.9e-64;
  Best Local Similarity
  Matches 217; Conservative 59; Mismatches 54;
                                                             Gaps
                                                 Indels
           4 MPLFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPT 63
QУ
             1 MPLFGKSQKSPVELVKSLKEAINALEAGDRKVEKAQEDVSKNLVSIKNMLYGSSDAEPPA 60
Db
          64 E-AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 122
 QY
             61 DYVVAQLSQELYNSNLLLLLIQNLHRIDFEGKKHVALIFNNVLRRQIGTRSPTVEYICTK 120
Db
         123 PHILFMLLKGYE--APQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIA 180
 Q7
             121 PEILFTLMAGYEDAHPEIALNSGTMLRECARYEALAKIMLHSDEFFKFFRYVEVSTFDIA 180
 Dic
         181 SDAFATFKDLLTRHKVLVADFLEQNYDTIF-EDYEKLLQSENYVTKRQSLKLLGELILDR 239
 QУ
             181 SDAFSTFKELLTRHKLLCAEFLDANYDKFFSQHYQRLLNSENYVTRRQSLKLLGELLLDR 240
 Db.
         240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
 QУ
```

```
241 HNFTVMTRYISEPENLKLMMNMLKEKSRNIQFEAFHVFKVFVANPNKPKPILDILLRNQT 300
Db
        300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK 333
Qу
            301 KLVDFLTNFHTDRSEDEQFNDEKAYLIKQIKELK 334
Db
RESULT 6
MO2M CAEEL
                              PRT; 338 AA.
                 STANDARD;
    MO2M CAEEL
ID
AC
    018211;
    16-OCT-2001 (Rel. 40, Created)
DT
    16-OCT-2001 (Rel. 40, Last sequence update)
TT
    16-OCT-2001 (Rel. 40, Last annotation update)
    Hypothetical MO25-like protein Y53C12A.4 in chromosome II.
DΤ
DΕ
    Y53C12A.4.
GN
    Caenorhabditis elegans.
    Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OS
OC.
    Rhabditidae; Peloderinae; Caenorhabditis.
OC.
    NCBI TaxID=6239;
OX
     [1]
PN
     SEQUENCE FROM N.A.
RP
     STRAIN=Bristol N2;
RC
     Kershaw J., Lennard N.;
RA:
    Submitted (SEP-1997) to the EMBL/GenBank/DDBJ databases.
RL
     ::- SIMILARITY: Belongs to the Mo25 family.
     ______
CC
     This SWISS-PROT entry is copyright. It is produced through a collaboration.
CC
     between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC
     the European Bioinformatics Institute. There are no restrictions on its
 CC
     use by non-profit institutions as long as its content is in no way
 CC
     modified and this statement is not removed. Usage by and for commercial
 CC
     entities requires a license agreement (See http://www.isb-sib.ch/announce/
 ĊC;
 CC
     or send an email to license@isb-sib.ch).
 CO
     CC
     EMBL; Z99277; CAB16486.1; -.
 ΌR
     PIR; T27129; T27129.
 ÐR
     WormPep; Y53C12A.4; CE14890.
 DR
     InterPro; IPR004892; Mo25.
 DR
     Pfam; PF03204; Mo25; 1.
 DR
     Hypothetical protein.
 KM
     SEQUENCE 338 AA; 39431 MW: 1D0C34A35D9116F5 CRC64;
 SQ
                       59.1%; Score 1006.5; DB 1; Length 338;
   Query Match
   Best Lecal Similarity 57.2%; Pred. No. 2.2e-57;
   Matches 191; Conservative 60; Mismatches 78; Indels
           5 PLFSKSHKNPAEIVKILKDNLAILEK-----QDKKTDKASEEVSKSLQAMKEILCGTNEK 59
             QУ
           4 PLFGKADKTPADVVKNLRDALLVIDRHGTNTSERKVEKAIEETAKMLALAKTFIYGSDAN 63
 dG
           60 EPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYI 119
             QУ
           64 EPNNEQVTQLAQEVYNANVLPMLIKHLHKFEFECKKDVASVFNNLLRRQIGTRSPTVEYL 123
 Db
          120 SAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDI 179
              QУ
```

```
124 AARPEILITLLLGYEQPDIALTCGSMLREAVRHEHLARIVLYSEYFQRFFVFVQSDVFDI 183
ďŒ
         180 ASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDR 239
             Qv
         184 ATDAFSTFKDLMTKHKNMCAEYLDNNYDRFFGQYSALTNSENYVTRRQSLKLLGELLLDR 243
Db
         240 HNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQP 299
             QУ
          244 HNFSTMNKYITSPENLKTVMELLRDKRRNIQYEAFHVFKIFVANPNKPRPITDILTRNRD 303
Db
          300 KLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK 333
              QУ
          304 KLVEFLTAFHNDRTNDEQFNDEKAYLIKQIQELR 337
Db
RESULT 7
YFV6 SCHPC
                                   PRT:
                    STANDARD;
     YFV6 SCHPO
ID
     Q9P7Q8;
AC
     16-OCT-2001 (Rel. 40, Created)
DΤ
     16-OCT-2001 (Rel. 40, Last sequence update)
היכד
     28-FEB-2003 (Rel. 41, Last annotation update)
 DT
     Hypothetical protein C1834.06c in chromosome I.
 DE
     SPAC1834.06C.
 GN
      Schizosaccharomyces pombe (Fission yeast).
      Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OS
 OC.
      Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC.
      schizosaccharomyces.
 OC.
      NCBI TaxID=4896;
 OX:
 RM
      [1]
      SEQUENCE FROM N.A.
 \mathbb{R}^{E}
      STRAIN=972;
 RC
      MEDLINE=21848401; PubMed=11859360;
      Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.
 RX.
      Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 R.A
      Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 RA
      Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
 RA
      Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 RA
      Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
 kΑ
      James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 RA
      Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 RI.
      Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
 RA
      Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
      Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
  RA
      Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
  RA
       Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
  RA
       Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
  RA
       Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
  RA
       Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
  RA
       Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
  RA.
       Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
  RA
       Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
  RA
       Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
  RA
       Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
  RA
       Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
  RA
       Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
  RA.
       Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
  RA
```

RA

```
"The genome sequence of Schizosaccharomyces pombe.";
RT
    Nature 415:871-880(2002).
RL
    -!- SIMILARITY: Belongs to the Mo25 family.
CC
    ______
CC
    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC
    the European Bioinformatics Institute. There are no restrictions on its
CC
    use by non-profit institutions as long as its content is in no way
CC
    modified and this statement is not removed. Usage by and for commercial
CC
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC
    or send an email to license@isb-sib.ch).
CC
    _____
CC
    EMBL; AL157734; CAB75774.1; -.
DR
    PIR; T50117; T50117.
DR
    GeneDB_SPombe; SPAC1834.06c; -.
DR
    InterPro; IPR004892; Mo25.
DR
    Pfam; PF03204; Mo25; 1.
DR
    Hypothetical protein.
KW
    SEQUENCE 329 AA; 38521 MW; 073DD0607A64C952 CRC64;
SQ
                      49.0%; Score 834.5; DB 1; Length 329;
  Query Match
  Best Local Similarity 51.5%; Pred. No. 1.9e-46;
 Matches 169; Conservative 63; Mismatches 93; Indels
                                                          Gaps / .
          5 LESKSHKNPAEIVKILKDNLAILE-KQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTE 64
OY
            1 LFNKRPKSTODVVRCLCDNLPKLEINNDKK--KSFEEVSKCLQNLRVSLCGTAEVEPDAD 61
         65 AVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAHPH 124
\mathbb{C}^{V}
             52 LVSDLSFQIYQSNLPFLLVRYLPKLEFESKKDTGLIFSALLRRHVASRYPTVDYMLAHPQ 121
Ďb
         125 ILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAF 184 W
Çy
            122 IFPVLVSYYRYQEVAFTAGSILRECSRHEALNEVLLNSRDFWTFFSLIQASSFDMASDAF 181
Db
         185 ATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQSENYVTKRQSLKLLGELILDRHNFAI 244
QY
            182 STFKSILLNHKSQVAEFISYHFDEFFKQYTVLLKSENYVTKRQSLKLLGEILLNRANRSV 241
Db
         245 MTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIEF 304
Qy
            242 MTRYISSAENLKLMMILLRDKSKNIQFEAFHVFKLFVANPEKSEEVIEILRRNKSKLISY 301.
Dh
         305 LSSFQKERTDDEQFADEKNYLIKQIRDL 332
 ÇУ
           · [[]: : [] : [] [] []: :: [] [] []
         302 LSAFHTDRKNDEQFNDERAFVIKQIERL 329
Db
 RESULT 8
 DE76 CHLPE
                 STANDARD;
                              PRT;
                                    321 AA.
     DE76 CHLPR
 ID
     OSXFY6;
 AC
     16-OCT-2001 (Rel. 40, Created)
 DT
     16-OCT-2001 (Rel. 40, Last sequence update)
 DT
     28-FEB-2003 (Rel. 41, Last annotation update)
 DT.
```

Degreening related gene dee76 protein.

DΕ

```
DEE76.
GN
    Chlorella protothecoides.
    Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OS
OC
    Chlorellaceae; Auxenochlorella.
OC
    NCBI TaxID=3075;
OX
    [1]
RN
    SEQUENCE FROM N.A.
RΡ
    STRAIN=ACC25;
RC
    MEDIJINE=20256472; PubMed=10798614;
    Hortensteiner S., Chinner J., Matile P., Thomas H., Donnison I.S.;
RX
    "Chlorophyll breakdown in Chlorella protothecoides: characterization
RA
RT
    of degreening and cloning of degreening-related genes.";
ŖТ
    Plant Mol. Biol. 42:439-450(2000).
RL
    -!- SIMILARITY: Belongs to the Mo25 family.
CC
    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC
    the European Bioinformatics Institute. There are no restrictions on its
CC
    use by non-profit institutions as long as its content is in no way
CC
    modified and this statement is not removed. Usage by and for commercial
CC
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC
CC
    or send an email to license@isb-sib.ch).
CC
     _____
CC
     EMBL; AJ238632; CAB42595.1; -.
DR
     InterPro; IPR004892; Mo25.
DR
     Pfam; PF03204; Mo25; 1.
SQ SEQUENCE 321 AA; 37262 MW; 918FD02964B09071 CRC64;
                       45.5%; Score 776; DB 1; Length 321;
  Query Match
  Best Local Similarity 52.0%; Pred. No. 9.8e-43;
  Matches 156; Conservative 56; Mismatches 84;
                                                         4; Gaps
                                                Indels
          33 DKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSGLLVTLIADLQLIDF 91
             19 ESKQDRVVEDISKAIMSIKEAIFGEDEQSSSKEHAQGIASEACRVGLVSDLVTYLTVLDF 78
 Db
          92 EGKKDVTQIFNNILRRQI--GTRSPTVEYISAHPHILFMLLKGYEAPQIALRCGIMLREC 149
             Qy
          79 ETRKDVVQIFCAIIRITLEDGGR-PGRDYVLAHPDVLSTLFYGYEDPEIALNCGQMFREC 137
 Db
          150 IRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTI 209
             OV
          138 IRHEDIAKFVLECNLFEELFEKLNVQSFEVASDAFATFKDLLTRHKQLVAAFLQENYEDF 197
 D_{ij}
          210 FEDYEKLLQSENYVTKRQSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNI 269
                QV
          193 FSQLDKLLTSDNYVTRRQSLKLLGELLLDRVNVKIMMQYVSDVNNLILMMNLLKDSSRSI 257
 D'o
          270 QFEAFHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQI 329
             ]]]]]]]]]]]]]]]]]]]]]
 QУ
          258 QFEAFHVFKVFVANPNKTKPVADILVNNKNKLLTYLEDFHNDR-DDEQFKEEKAVIIKEI 316
```

RESULT 9
MO2N\_ARATH
ID MO2N\_ARATH STANDARD; PRT; 343 AA.
AC Q9FGK3;

```
16-OCT-2001 (Rel. 40, Created)
DT
    16-OCT-2001 (Rel. 40, Last sequence update)
DT
    16-OCT-2001 (Rel. 40, Last annotation update)
DT
    Hypothetical MO25-like protein At5g47540.
DΕ
    AT5G47540 OR MNJ7.13.
GN
    Arabidopsis thaliana (Mouse-ear cress).
OS
    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC
    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC
    eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OC
    NCBI TaxID=3702;
OX
RN
    [1]
    SEQUENCE FROM N.A.
RР
    STRAIN=cv. Columbia;
RC
    Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA
    Tabata S.;
RA
    "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RT
    Submitted (APR-1999) to the EMBL/GenBank/DDBJ databases.
RL
    -!- SIMILARITY: Belongs to the Mo25 family.
CC
    \mathbb{C}\mathbb{C}
    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC
    the European Bioinformatics Institute. There are no restrictions on its
CC
    use by non-profit institutions as long as its content is in no way
CC
    modified and this statement is not removed. Usage by and for commercial
CC
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC
    or send an email to license@isb-sib.ch).
CC
    -----
CC
    EMBL; AB025628; BAB09080.1; -.
DR.
    InterPro: TPR004892; Mo25.
-DR.
    Pfam; PF03204; Mo25; 1.
DR
    Hypothetical protein.
ΧW
    SEQUENCE 343 AA; 39457 MW; 46950D6A9A82FBB5 CRC64;
SO
                      42.7%; Score 728; DB 1; Length 343;
  Query Match
  Pest Local Similarity 43.2%; Pred. No. 1.2e-39;
  Matches 147; Conservative 79; Mismatches 100; Indels
           6 LFSKSHKNPAEIVKILKDNLAILEK------QDKKTDKASEEVSKSLQAMKEILCGTNE 58
Qy
            4 LFKSKPRTPADLVRQTRDLLLFSDRSTSLPDLRDSKRDEKMAELSRNIRDMKSILYGNSE 63
מנו
          59 KEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEY 118
QУ
             54 AEPVAEACAQLTQEFFKEDTLRLLITCLPKLNLETRKDATQVVANLQRQQVNSRLIASDY 123
Db
         119 ISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFD 178
QY
             124 LEANIDLMDVLIEGFENTDMALHYGAMFRECIRHQIVAKYVLESDHVKKFFDYIQLPNFD 183
Db
         179 IASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDY-EKLLQSENYVTKRQSLKLLGELIL 237
Q\v
             184 IAADAAATFKELLTRHKSTVAEFLTKNEDWFFADYNSKLLESSNYITRRQAIKLLGDILL 243
D_{\mathcal{O}}
         238 DRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKN 297
Qy.
             244 DRSNSAVMTKYVSSRDNLRILMNLLRESSKSIQIEAFHVFKLFAANQNKPADIVNILVAN 303
Db
```

RESULT 10 MO2M ARATH 343 AA. PRT; STANDARD; MO2M ARATH ΞD Q9M0M4; 023570; AC16-OCT-2001 (Rel. 40, Created) DT16-OCT-2001 (Rel. 40, Last sequence update) DT15-SEP-2003 (Rel. 42, Last annotation update) DT Hypothetical MO25-like protein At4g17270. DΕ AT4G17270 OR DL4670W. GNArabidopsis thaliana (Mouse-ear cress). Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; OS Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; OC OC eurosids II; Brassicales; Brassicaceae; Arabidopsis. OC NCBI TaxID=3702; ΟX RN[1] SEQUENCE FROM N.A. RPSTRAIN=cv. Columbia; RC MEDLINE=98121113; PubMed=9461215; Bevan M., Bancroft I., Bent E., Love K., Goodman H.M., Dean C.,  $\mathcal{P}Z$ Bergkamp R., Dirkse W., van Staveren M., Stiekema W., Drost L., RARidley P., Hudson S.-A., Patel K., Murphy G., Piffanelli P., RAWedler H., Wedler E., Wambutt R., Weitzenegger T., Pohl T., Terryn N.,  $\mathbb{R}A$ Gielen J., Villarroel R., De Clercq R., van Montagu M., Lecharny A.,  $\mathbb{R}\mathbb{R}$ Aubourg S., Gy I., Kreis M., Lao N., Kavanagh T., Hempel S., 22Notter P., Entian K.-D., Rieger M., Schaefer M., Funk B., ř.  $\mathbb{G}[X]$ 25 Mueller Auer S., Silvey M., James R., Monfort A., Pons A., K.B 1870 Fuigdomenech P., Douka A., Voukelatou E., Milioni D., Hatzopoulos P., 174. Dirayandi E., Obermaier B., Hilbert H., Duesterhoeft A., Moores T.,  $T_{i} N$ Jones J.D.G., Eneva T., Palme K., Benes V., Rechmann S., Ansorge W., الماراتين الماراتين Cooke R., Berger C., Delseny M., Voet M., Volckaert G., Mewes H.-W., RA Klosterman S., Schueller C., Chalwatzis N.; KLÀ "Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of RA $\mathbb{R}^{T}$ Arabidopsis thaliana."; Nature 391:485-488(1998). 2011  $\mathbf{F}_{-}$ SEQUENCE FROM N.A. ŖĐ. STRAIN=cv. Columbia; PC. MEDLINE=20083488; PubMed=10617198; Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G., RX. Pohl To, Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N., RAHarris B., Ansorge W., Brandt P., Grivell L.A., Rieger M., ŔΑ Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M., KΑ Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T., ·RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I., PAVos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P., RA RAHangham S.-A., McCullagh B., Bilham L., Robben J., Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F., Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E., A $\mathcal{A}$ Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M., RAHolzer E., Brandt A., Peters S., van Staveren M., Dirkse W., RΑ Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P., RABerneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H., RARA

```
De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
    Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA
    Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA
    Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA
    Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA
    Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA
    Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA
     Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA
     Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA
     Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA
     Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA
     Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA
     Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA
     Heijnen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA
RA
     Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
     Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA
     Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA
     Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA
     Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA
     Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA
     Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA
     Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA
     Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA
     Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
PA
     Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA
     Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA
     Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA
     Granat S., Shondy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
\mathbb{R}T_{i}
     Chen E., Marra M., Martienssen R., McCombie W.R.;
\mathbb{H}\mathbf{A}
     "Sequence and analysis of chromosome 4 of the plant Arabidopsis,
 £22.
\mathbf{F}^{*}\mathcal{F}
     thaliana.";
PT
      Nature 402:769-777(1999).
 RL.
 3N
      [3]
      SEQUENCE FROM N.A.
 PP
     STRAIN=cv. Columbia;
      Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 .EC
      "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 \Sigma \Delta
 RT
      SSP consortium (Salk/Stanford/PGEC).";
 RT
      Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
 RL
      -!- SIMILARITY: Belongs to the Mo25 family.
      -4- CAUTION: Ref.1 sequence differs from that shown due to erroneous
 CC
 CC
        gene model prediction.
      CC
      This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC
      between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC
      the European Bioinformatics Institute. There are no restrictions on its
 CC
      onse by non-profit institutions as long as its content is in no way
 CC
      modified and this statement is not removed. Usage by and for commercial
 CC.
      entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC
 CC
      or send an email to license@isb-sib.ch).
 CC
      ĊC
      EMBL; 797343; CAB10508.1; ALT_SEQ.
 DŔ
      EMBL; AL161546; CAB78730.1; -.
  DR
      EMBL; AF380659; AAK55740.1; -.
  DR
      InterPro; IPR004892; Mo25.
  DR.
      Pfam; PF03204; Mo25; 1.
  DR
      Hypothetical protein.
  KW
```

```
SEQUENCE 343 AA; 39650 MW; D340B49A4924B7D1 CRC64;
SQ
                       42.0%; Score 716.5; DB 1; Length 343;
 Query Match
                      42.9%; Pred. No. 6.5e-39;
 Best Local Similarity
 Matches 144; Conservative 78; Mismatches 105; Indels
                                                                    3;
                                                             Gaps
          6 LFSKSHKNPAEIVKII,KDNLAILEK------QDKKTDKASEEVSKSLQAMKEILCGTNE 58
            Qу
          4 LFKSKPRTPADIVRQTRDLLLYADRSNSFPDLRESKREEKMVELSKSTRDLKLILYGNSE 63
Db
          59 KEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEY 118
             QУ
          54 AEPVAEACAQLTQEFFKADTLRRLLTSLPNLNLEARKDATQVVANLQRQQVNSRL1AADY 123
Db
         119 ISAHPHILFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFD 178
             ÕΆ
         124 LESNIDLMDFLVDGFENTDMALHYGTMFRECIRHQIVAKYVLDSEHVKKFFYYIQLPNFD 183
d\mathbf{G}
         179 TASDAFATFKOLLTRHKVLVADFLEQNYDTIFEDY-EKLLQSENYVTKRQSLKLLGELIL 237
             HENCHMENTH HEN ELL EN THE HEITENHEEL
Qу
         184 IAADAAATFKELLTRHKSTVAEFLIKNEDWFFADYNSKLLESTNYITRRQAIKLLGDILL 243
 ďŒ
         238 DRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLKN 297 - CAL
             QУ
          244 DRSNSAVMTKYVSSMDNLRILMNLLRESSKTIQIEAFHVFKLFVANQNKPSDIANILVAN 303
 Db
          303 OPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK 333
             ATT: 1: 1: 1: 1 ATT: 11
 QT
          304 RNKLLRLLADIKPDK-EDERFDADKAQVVREIANLK 338
 DO
 RESULT 11
 HYMA EMENI
                                       384 AA.
                                PRT;
                   STANDARD;
      HYMA EMENI
 ID
      060032;
 AC
      16-OCT-2001 (Rel. 40, Created)
      16-OCT-2001 (Rel. 40, Last sequence update)
 DT
      16-OCT-2001 (Rel. 40, Last annotation update)
 ÐΤ
 DT
      Conidiophore development protein hymA.
 DΕ
      Emericella nidulans (Aspergillus nidulans).
 GN
      Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
      Eurotiales; Trichocomaceae; Emericella.
  OC.
      NCBI TaxID=162425;
  OX
      [1]
  RN
      SEQUENCE FROM N.A.
  2.D
      MEDLINE=99126010; PubMed=9928930;
  RX
      Karos M., Fischer R.;
      "Molecular characterization of HymA, an evolutionarily highly
  RA
      conserved and highly expressed protein of Aspergillus nidulans.";
  RT
  RT
      Mol. Gen. Genet. 260:510-521(1999).
      -!- FUNCTION: Required for conidiophore development.
  RL
  CC
      -!- SUBCELLULAR LOCATION: Cytoplasmic.
  CC
       -!- SIMILARITY: Belongs to the Mo25 family.
  CC
      This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC
       between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC
  CC
```

```
the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
CC
    modified and this statement is not removed. Usage by and for commercial
CC
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC
CC
    or send an email to license@isb-sib.ch).
CC
    ______
CC
    EMBL; AJ001157; CAA04556.1; -.
DR
    InterPro: IPR004892; Mo25.
DR
    Pfam; PF03204; Mo25; 1.
DR
              384 AA; 44392 MW; 2E203D0D110C5FD6 CRC64;
    SEQUENCE
SO
                       39.1%; Score 666; DB 1; Length 384;
  Ouery Match
  Best Local Similarity 39.8%; Pred. No. 1.2e-35;
  Matches 140; Conservative 68; Mismatches 114; Indels
                                                          30; Gaps
          12 KNPAEIVKILKDNLAILEKODKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQ 71
             QУ
          11 RQPSDVVRSIKDLLLRL-REPSTASKVEDELAKQLSQMKLMVQGTQELEASTDQVHALVO 69
Db
          72 ELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILR----RQIGTRSPTVEYI-SAHPHIL 126
             Çγ
                                                      1: ||
                                                               | | | :
          70 AMLHEDLLYELAVALHNLPFEARKDTQTIFSHILRFKPPHGNSPDPPVISYIVHNRPEII 129
Dò
 * 3
          137 FMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQ---
 ÛУ
               4 STH (1 1: 11 STH SEE, ST 1 HE
         189 TELCRGYEHSQSAMPCGTILREALKFDVIAAIILYDQSKEGEPAIRLTEVQPNV9QRGTG 189
 i)];
          165 -FRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEK-LLQSENY 222
              190 VFWRFFHWIDRGTFELSADAFTTFREILTRHKSLVTGYLATNFDYFFAQFNTFLVQSESY 249
 7323
          223 VIKRQSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVA 282
             Q_{\overline{I}}
          250 VTKROSIKLLGEILLDRANYSVMMRYVESGENLKLCMKLLRDDRKMVQYEGFHVFKVFVA 309
 勁
          283 SPHKTQPIVEILLKNQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLKK 334
              :| |: : : ||: |: :|: || | | ::||||:|| |||:|::|| | | |
 ŲΨ
          310 NPDKSVAVQRILINNRDRLLRFLPKFLEDRTDDDQFTDEKSFLVRQIELLPK 361
 Ľb
 RESULT 12
 MO2L ARATH
                                       348 AA.
                   STANDARD;
                                 PRT:
      MO2L ARATH
 ID
      Q9ZQ77;
 AC
      16-OCT-2001 (Rel. 40, Created)
 TC
      16-OCT-2001 (Rel. 40, Last sequence update)
 TYT
      16-OCT-2001 (Rel. 40, Last annotation update)
 DT
      Hypothetical MO25-like protein At2g03410.
 DE
      AT2G03410 OR T4M8.16.
 GN
      Arabidopsis thaliana (Mouse-ear cress).
      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OS
      Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC.
 OC
      eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC.
      NCBI TaxID=3702;
 OX
  RM
      [1]
      SEQUENCE FROM N.A.
  RP
      STRAIN=cv. Columbia;
  RC
```

```
MEDLINE=20083487; PubMed=10617197;
    Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RX
RA
    Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA
    Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA
    Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA
    Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA
    Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
    Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA
RA
    Venter J.C.;
RA
    "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT
    thaliana.";
RT
    Nature 402:761-768(1999).
RL
    -!- SIMILARITY: Belongs to the Mo25 family.
CC
    ______
CC
    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC
    the European Bioinformatics Institute. There are no restrictions on its
CC
    use by non-profit institutions as long as its content is in no way
CC
    modified and this statement is not removed. Usage by and for commercial
CC.
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC
    or send an email to license@isb-sib.ch).
CC
     CC
    EMBL; AC006284; AAD17435.1; -.
    PIR; B84448; B84448.
DR
     InterPro; IPR004892; Mo25.
DR
    Pfam; PF03204; Mo25; 1.
DR
    Hypothetical protein.
\sim
    SEQUENCE 348 AA; 40000 MW; AB1D92EA2E2B900E CRC64;
50 ·
                       37.1%; Score 632; DB 1; Length 348;
  Query Match
  Best Local Similarity 38.7%; Pred. No. 1.6e-33;
  Matches 133; Conservative 80; Mismatches 117; Indels 14; Gaps 35
           6 GFSKSHKNPAEIVKILKDNLAILEKQDKKTDKASE-----EVSKSLQAMKEILCGTNE 58
            4 LFKNKSRLPGEIVRQTRDLIALAESEEEETDARNSKRLGICAELCRNIRDLKSILYGNGE 63
          59 KEPPTEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEY 118
              QV
          64 AEPVPEACLLLTQEFFRADTLRPLIKSIPKLDLEARKDATQIVANLQKQQVEFRLVASEY 123
         119 ISAHPHILFMLLKGYEAP-QIALRCGIMLRECTRHEPLAKTILFSNQFRDFFKYVELSTF 177
             ÇΨ
         124 LESNLDVIDSLVEGIDHDHELALHYTGMLKECVRHQVVAKYILESKNLEKFFDYVQLPYF 183
         178 DIASDAFATFKOLLTRHKVLVADFLEQNYDTIFEDYE-KLLQSENYVTKRQSLKLLGELI 236
             THERE THE HELL THE THE THE THE THEFT
 \Omega V
         184 DVATDASKIFRELLTRHKSTVAEYLAKNYEWFFAEYNTKLLEKGSYFTKRQASKLLGDVL 243
 Db
          237 LDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVEILLK 296
 QУ
             244 MDRSNSGVMVKYVSSLDNLRIMMNLLREPTKNIQLEAFHIFKLFVANENKPEDIVAILVA 303
 Db
          297 NQPKLIEFLSSFQKERTDDEQFADEKNYLIKQIRDLK----KTA 336
 Qу
             1: 1:: : : 1:: : 1 : : : : 1 : : : : 1
          304 NRTKILRLFADLKPEK-EDVGFETDKALVMNEIATLSLLDIKTA 346
 UC
```

```
RESULT 13
HYM1 YEAST
                                          399 AA.
                                   PRT:
                    STANDARD;
    HYM1_YEAST
ID
     P32464;
AC
     01-OCT-1993 (Rel. 27, Created)
DT
     01-OCT-1993 (Rel. 27, Last sequence update)
DТ
     16-OCT-2001 (Rel. 40, Last annotation update)
DT
     HYM1 protein.
DΕ
     HYM1 OR YKL189W.
GN
     Saccharomyces cerevisiae (Baker's yeast).
OS
     Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC
     Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OC
     NCBI 'TaxID=4932;
NΟ
RN
     [1]
     SEQUENCE FROM N.A.
RP -
     STRAIN=GRF88;
RC
     MEDLINE=93348778; PubMed=8394042;
RX
     Cheret G., Mattheakis L.C., Sor F.;
R\Lambda
     "DNA sequence analysis of the YCN2 region of chromosome XI in
RT
     Saccharomyces cerevisiae.";
RT
     Yeast 9:661-667(1993).
RL
RN
     [2]
     SEQUENCE FROM N.A.
RP
     MEDLINE=94205264; PubMed=8154185;
\mathbf{x}
     Wiemann S., Voss H., Schwager C., Rupp T., Stegemann J.,
RA
     Zimmermann J., Grothues D., Sensen C., Erfle H., Hewitt N.,
I.F.
     Banrevi A., Ansorge W.;
EA
     "Sequencing and analysis of 51.6 kilobases on the left arm of
\mathbb{R}\mathbb{Z}
     chromosome XI from Saccharomyces cerevisiae reveals 23 open reading
2T
      frames including the FAS1 gene.";
ŔΤ
      Yeast 9:1343-1348(1993)
RL
22N
      SEQUENCE FROM N.A.
ŔР
     Maia e Silva A., Bossier P., Vilela C., Fernandes L., Soares H.,
RA
      Guerreiro P., Rodrigues-Pousada C.;
 RA
      Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.
 R\mathbf{L}
 RN
 RΡ
      GENE NAME.
      MEDLINE=20157038; PubMed=10655212;
 RX
      Dorland S., Deegenaars M.L., Stillman D.J.;
      *Roles for the Saccharomyces cerevisiae SDS3, CBK1 and HYM1 genes in
 RA
 RT
      transcriptional repression by SIN3.";
 \mathbb{R}^{\mathbf{T}}
      Genetics 154:573-586(2000).
 RL
      -!- SIMILARITY: Belongs to the Mo25 family.
 CC
      ĊĊ
      This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC
      between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC
      the European Bioinformatics Institute. There are no restrictions on its
 CC
      use by non-profit institutions as long as its content is in no way
 CC
      modified and this statement is not removed. Usage by and for commercial
      entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC
 CC
      or send an email to license@isb-sib.ch).
 CC
 CC
      EMBL; X69765; CAA49422.1; -.
 DR
      EMBL; X74151; CAA52249.1; -.
 DR
      EMBL; Z28189; CAA82032.1; -.
 DR
```

```
PIR; S34681; S34681.
DR
         SGD; S0001672; HYM1.
DR
         GO; GO:0005622; C:intracellular; IDA.
DK
         GO; GO:0016564; F:transcriptional repressor activity; IMP.
         GO; GO:0007109; P:cytokinesis, completion of separation; IMP.
DR
DR
         GO; GO:0008360; P:regulation of cell shape; IGI.
DR
         InterPro; IPR004892; Mo25.
DR
         Pfain; PF03204; Mo25; 1.
                          399 AA; 45853 MW; F48860754C892BA9 CRC64;
DR
         SEQUENCE
SQ
                                               28.5%; Score 485; DB 1; Length 399;
    Ouery Match
                                                33.0%; Pred. No. 4.3e-24;
    Best Local Similarity
    Matches 113; Conservative 75; Mismatches 138; Indels
                                                                                                                              Gaps
                       7 FSKSHKNPAEIVKILKDNLAILEK----QDKKTDKASEEVSKSLQAMKEILCGTNEKEPP 62
                                                                  ÇΥ
                           . |: | |:: ::: : |
                     16 WKKNPKTPSDYARLIIEQLNKFSSPSLTQDNKR-KVQEECTKYLIGTKHFIVGDTDPHPT 74
 Db
                     53 TEAVAQLAQELYSSGLLVTLIADLQLIDFEGKKDVTQIFNNILRRQIGTRSPTVEYISAH 1224
                            floring the state of the state 
 Qy
                     75 PEAIDELYTAMHRADVFYELLLHFVDLEFEARRECMLIFSICLGYSKDNKFVTVDYLVSQ 134
 Tib
                   230 RHILFMLLKGYE-----APQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELS 175 (2)
                                                                   QV
                           1 :::::
                    PKTISLMLRTAEVALQOKGCQDIFLTVGNMIIECIKYEQLCRIILKDPQLWKFFEFAKLG 194
 J.b
                    176 TFDIASDAFATFKDLLTRHKVLVA-DFL--EQNYDTIFEDYEKLLQSENYVTKROSLKLL 232
                                                        QУ
                    195 NFEISTESLQILSAAFTAHPKLVSKEFFSNEINIIRFIKCINKLMAHGSYVTKRQSTKLL 254
                    333 GELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFKVFVASPHKTQPIVE 292
                               A.
                    255 ASLIVIRSNNALMNIYINSPENLKLIMTLMTDKSKNLQLEAFNVFKVMVANPRKSKPVFD 314
  3.2h
                    293 ILLKNOPKLIEFLSSFOKERTDDEQFADEKNYLIKQIRDLKK 334
                            897
                    315 ILVKNRDKLLTYFKTFGLD-SQDSTFLDEREFIVQEIDSLPR 355
   Db
   RESULT 14
   MO2L CAEEL
                                                                               339 AA.
                                                                   PRT;
                                       STANDARD;
            MO2L CAEEL
            Q9TZM2;
            16-OCT-2001 (Rel. 40, Created)
   TG
            16-OCT-2001 (Rel. 40, Last sequence update)
   DT
            15-OCT-2001 (Rel. 40, Last annotation update)
            Hypothetical MO25-like protein T27C10.3 in chromosome I.
   TCI
   DΕ
            T27C10.3.
   CN
             Caenorhabditis elegans.
             Mukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
   OS
    00
             Phabditidae; Peloderinae; Caenorhabditis.
    OC.
             MCBI_TaxID=6239;
    OX
    8N
             SEQUENCE FROM N.A.
    RΡ
             STRAIN=Bristol N2;
    RC
             Zhu H.J., Graves T., Hawkins M.;
             Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
    RA
    RL
```

```
-!- SIMILARITY: Belongs to the Mo25 family.
CC
    ______
    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC
    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC
CC
    the European Bioinformatics Institute. There are no restrictions on its
    use by non-profit institutions as long as its content is in no way
CC
    modified and this statement is not removed. Usage by and for commercial
CC
    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC
CC
    or send an email to license@isb-sib.ch).
CC
    ______
CC
    EMBL; AF098504; AAC67411.1; -.
DR
    PIR; T33477; T33477.
DR
    WormPep; T27C10.3; CE19605.
DR
    InterPro; IPR004892; Mo25.
DR
    Pfam; PF03204; Mo25; 1.
DR
    Hypothetical protein.
KW.
    SEQUENCE 339 AA; 40232 MW; E7DA45CA33F2947E CRC64;
30
                       8.4%; Score 143.5; DB 1; Length 339;
  Query Match
  Rest Local Similarity 19.3%; Pred. No. 0.02;
  Matches 38; Conservative 50; Mismatches 76; Indels
         159 IDFSNQFRDFFKYVELSTFDIASDAFATFKDLLTRHKVLVADFLEQNYDTIFEDYEKLLQ 218
             QΥ
         100 LMNTNKFRD------FINVIQGTFDTLQIIFFTNHESANNFIKNNLPRFMQTLHKLIA 150
Ob.
         219 SENYVTKRQSLKLLGELILDRHNFAIMTKYISKPENLKLMMNLLRDKSPNIQFEAFHVFK 278
              Q'Z
         151 CSNFFIQAKSFKFLNELFTAQTNYETRSLWMAEPAFIKLVVLAIQSNKHAVRSRAVSILE 210
DD:
         279 VFVASPHKTQPIVEILLKNQPKLIEFL------SSFQKERTDDEQFAD ----- 320 %
         QУ
         211 IFIRNPRNSPEVHEFIGRNRNVLIAFFFNSAPIHYYQGSPNEKE---DAQYARMAYKLLN 267
. 15,2
         321 --- EKNYLIKQIRDLKK 334
 ŨΫ
                :: : : : : : : : : : :
         268 WDMQRPFTQEQLQDFEE 284
 RESULT 15
 AKA9 HUMAN
                 STANDARD; PRT; 3911 AA.
    AKA9 HUMAN
    Q99996; O14869; O43355; O94895; Q9UQH3; Q9UQQ4; Q9Y6B8; Q9Y6Y2;
     16-OCT-2001 (Rel. 40, Created)
     16-OCT-2001 (Rel. 40, Last sequence update)
 TC
     15-SEP-2003 (Rel. 42, Last annotation update)
 DT
     A-kinase anchor protein 9 (Protein kinase A anchoring protein 9)
     (PRKA9) (A-kinase anchor protein 450 kDa) (AKAP 450) (A-kinase anchor
 DE
     protein 350 kDa) (AKAP 350) (hgAKAP 350) (AKAP 120 like protein)
 DΕ
     (Hyperion protein) (Yotiao protein) (Centrosome- and Golgi-localized
 DE
 DΕ
     PKN-associated protein) (CG-NAP).
 DΕ
     AKAP9 OR AKAP450 OR AKAP350 OR KIAA0803.
 GN
     Homo sapiens (Human).
     Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
     Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC
     NCEI_TaxID=9606;
 OX
      [1]
 RN
```

```
SEQUENCE FROM N.A. (ISOFORM 4).
RP
     TISSUE=Brain;
RC
     MEDLINE=98151389; PubMed=9482789;
     Lin J.W., Wyszynski M., Madhavan R., Sealock R., Kim J.U., Sheng M.;
RX
     "Yotiao, a novel protein of neuromuscular junction and brain that
RA
     interacts with specific splice variants of NMDA receptor subunit
RT
RT
RT
     J. Neurosci. 18:2017-2027(1998).
RL
     SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANT GLN-1347 INS.
RN
RΡ
     MEDLINE=99219864; PubMed=10202149;
     Witczak O., Skaalhegg B.S., Keryer G., Bornens M., Tasken K.,
XX
RA
     Jahnsen T., Oerstavik S.;
     "Cloning and characterization of a cDNA encoding an A-kinase anchoring
RA
RT
     protein located in the centrosome, AKAP450.";
RT
     EMBO J. 18:1858-1868(1999).
RL.
RN
     SEQUENCE FROM N.A. (ISOFORM 3).
RΡ
      TISSUE=Brain;
RC.
     MEDI.INE=99287934; PubMed=10358086;
     Takahashi M., Shibata H., Shimakawa M., Miyamoto M., Mukai H., Ono Y.;
RX
      "Characterization of a novel giant scaffolding protein, CG-NAP, that
RΑ
      anchors multiple signaling enzymes to centrosome and the Golgi
RT
TS
      apparatus.";
RT
      J. Biol. Chem. 274:17267-17274(1999).
 27.
 \mathbb{R}\mathbb{N}
      14
      SEQUENCE FROM N.A. (ISOFORM 1).
 \mathbb{RP}
      Kemmner W.A., Deiss S., Schwarz U.;
 PA
      "Cloning of Hyperion.";
 TS
      Submitted (AUG-1998) to the EMBL/GenBank/DDBJ databases.
 101
      (5)
 RM \sim
      SEQUENCE OF 323-3911 FROM N.A. (ISOFORM 2).
 PP
      TISSUE=Gastric parietal cell;
 {\rm R}{\mathbb C}
      MEDLINE=99115654; PubMed=9915845;
      Schmidt P.H., Dransfield D.T., Claudio J.O., Hawley R.G.,
 RX
 RA
      Trotter K.W., Milgram S.L., Goldenring J.R.;
      "AKAP350, a multiply spliced protein kinase A-anchoring protein
 RA
 \mathbf{r}
      associated with centrosomes.";
 RT
      J. Biol. Chem. 274:3055-3066(1999).
 RL
 RM
      SEQUENCE OF 1802-3876 FROM N.A. (ISOFORM 5).
 RP
       TISSUE=Lymphoblast;
 RC
      Hinds K., Sutterer C., Becker M., Hawkins M.;
      Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
 RA
 RL
  PM
       SEQUENCE OF 2157-3911 FROM N.A. (ISOFORM 6).
  ΣĎ
       TISSUE=Lung;
  RС
       Milgram S.L., Goldenring J.R., Schmidt P.H.;
       "AKAP350: A multiply spliced family of proteins with centrosomal
  RA
  RT.
       association.";
       Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
  \Sigma^{\mathrm{T}}
  RL
  RN
       SEQUENCE OF 2212-3911 FROM N.A. (ISOFORM 2/3).
  RP
       TISSUE=Brain;
  RC
       MEDLINE=99087487; PubMed=9872452;
       Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
  RX
  RA
       Tanaka A., Kotani H., Nomura N., Ohara O.;
  RA
```

```
"Prediction of the coding sequences of unidentified human genes. XI.
    The complete sequences of 100 new cDNA clones from brain which code
RT
RT
     for large proteins in vitro.";
RT
     DNA Res. 5:277-286(1998).
RL
     [9]
RN
    SEQUENCE OF 17-1800 FROM N.A.
RΡ
     Wu X., Graves T., Bradshaw H.;
     Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.
RЯ
     -!- FUNCTION: BINDS TO TYPE II REGULATORY SUBUNITS OF PROTEIN KINASE
RL
         A. SCAFFOLDING PROTEIN THAT ASSEMBLES SEVERAL PROTEIN KINASES AND
CC
         PHOSPHATASES ON CENTROSOME AND GOLGI APPARATUS WHERE PHYSIOLOGICAL
CC
         EVENTS CAN BE REGULATED BY PHOSPHORYLATION STATE OF PROTEIN
CC
         SUBSTRATES. ISOFORM 4/YOTIAO IS ASSOCIATED WITH THE N-METHYL-D-
CC
         ASPARTATE RECEPTOR AND IS SPECIFICALLY FOUND IN THE NEUROMUSCULAR
CC
         JUNCTION (NMJ) AS WELL AS IN NEURONAL SYNAPSES EXPLAINING THAT ITS
CC
         ROLE MAY BE TO ORGANIZE POSTSYNAPTIC SPECIALIZATIONS.
CC
     -!- SUBUNIT: INTERACTS WITH THE REGULATORY REGION OF PROTEIN KINASE N
CC
         (PKN), PROTEIN PHOSPHATASE 2A (PP2A), PROTEIN PHOSPHATASE 1 (PP1)
СC
         AND THE IMMATURE NON-PHOSPHORYLATED FORM OF PKC EPSILON.
CC
     - :- SUBCELLULAR LOCATION: CENTROSOMAL IN MANY CELL TYPES AND
CC
 CC
         CYTOPLASMIC IN PARIETAL CELLS.
CC
      -!- ALTERNATIVE PRODUCTS:
         Event=Alternative splicing; Named isoforms=6;
 CC
 CC -
         Name=1;
 CC
           Toold=Q99996-1; Sequence=Displayed;
 CC
          Wame=2;
          IsoId=Q99996-2; Sequence=VSP_004102, VSP_004107;
 CC
 CC.
          Name=3: Synonyms=CG-NAP;
          IsoId=Q99996-3; Sequence=VSP_004102, VSP_004105, VSP_004107;
 CC
 CC
          mame=4; Synonyms=Yotiao;
          IsoId=Q99996-4; Sequence=VSP_004103, VSP_004104;
 CC
 Mame=5;
         IsoId=Q99996-5; Sequence=VSP_004108;
 er e ere
Sinhaa
          Name=6; Synonyms=AKAP350;
        IsoId=Q99996-6; Sequence=VSP_004106, VSP_004107, VSP_004109;
 · CC
      -!- TISSUE SPECIFICITY: WIDELY EXPRESSED. ISOFORM 4/YOTIAO IS HIGHLY
 GC
 CC
          EXPRESSED IN SKELETAL MUSCLE AND IN PANCREAS.
      -!- DOMAIN: RII BINDING SITE, PREDICTED TO FORM AN AMPHIPATHIC HELIX,
 CC
          COULD PARTICIPATE IN PROTEIN-PROTEIN INTERACTIONS WITH A
 CC
 CC
          COMPLEMENTARY SURFACE ON THE R-SUBUNIT DIMER.
      -!- CAUTION: REF.6 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO TWO
 CC
  CC
          FRAMESHIFTS IN POSITIONS 3782 AND 3811.
       -!- CAUTION: REF.9 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FOUR
  CO
          FRAMESHIFTS IN POSITIONS 29, 1653, 1699 AND 1735.
  CC
       CC
      This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC
      between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC
       the European Bioinformatics Institute. There are no restrictions on its
  CC
       use by non-profit institutions as long as its content is in no way
  ·CC
       modified and this statement is not removed. Usage by and for commercial.
  CC
       entities requires a license agreement (See http://www.isb-sib.ch/announce/
  CC
  CC
       or send an email to license@isb-sib.ch).
  CC
  CC
       EMBL; AJ131693; CAB40713.1; -.
  DR
       EMBL; AB019691; BAA78718.1; -.
  DR
       EMBL; AJ010770; CAA09361.1; -.
  DR
       EMBL; AF026245; AAB86384.1; -.
  DR
```

```
EMBL; AF083037; AAD22767.1; -.
DR
     EMBL; AC004013; AAB96867.1; ALT_FRAME.
DR
     EMBL; AF091711; AAD39719.1; -.
DR
     EMBL; AB018346; BAA34523.1; -.
DR
     EMBI; AC000066; AAC60380.1; ALT_FRAME.
DR
      Genew; HGNC:379; AKAP9.
DR
      MIM; 604001; -.
DR
      GO; GO:0005813; C:centrosome; TAS.
DR
      GO; GO:0005856; C:cytoskeleton; TAS.
      GO; GO:0004973; F:N-methyl-D-aspartate receptor-associated pr. . .; TAS.
DR
DR
      GO; GO:0005515; F:protein binding activity; TAS.
DR
      GO; GO:0007165; P:signal transduction; TAS.
DR
      GO; GO:0006832; P:small molecule transport; TAS.
DR
      GO; GO:0007268; P:synaptic transmission; TAS.
DR
      Coiled coil; Alternative splicing; Polymorphism.
                                        PKA-RII SUBUNIT BINDING DOMAIN.
KW
                           2567
                   2554
      DOMAIN
FT
                                        COILED COIL (POTENTIAL).
                            914
                    164
\overline{r}T
      MIAMOG
                                        COILED COIL (POTENTIAL).
                           1022
                    944
      DOMAIN
 FT
                                        COILED COIL (POTENTIAL).
                           1185
                   1100
      DOMAIN
ΥT
                                        COILED COIL (PCTENTIAL).
                           1280
                   1253
 FT
      DOMAIN
                                        COILED COIL (POTENTIAL).
                           1392
                   1336
 FT
      DOMAIN
                                        COILED COIL (POTENTIAL).
                           1459
                   1434
      DOMAIN
 FT
                                        COILED COIL (POTENTIAL).
                           1659
                   1585
      DOMAIN
 FT
                                        COILED COIL (POTENTIAL).
                           2455
       DOMAIN
                   1857
 FT
                                         COILED COIL (POTENTIAL).
                           2561
                   2544
 P^{\prime \Gamma}
       DOMAIN
                                         COILED COIL (POTENTIAL).
                            2776
                    2603
 FT
       MIAMOC
                                         COILED COIL (POTENTIAL).
                    3065
                           3092
 3/12
       DOMAIN
                                         COILED COIL (POTENTIAL).
                            3470
                    3124
 202
       DOMAIN
                                         COILED COIL (POTENTIAL).
                            3689
                    3587
 \mathbb{R}^{r} \Gamma
       DOMAIN
                                         POLY-LEU.
                            3730
                    3726
 aprilia.
       DOMAIN
                                         GLN-RICH.
                             292
                     203
 <u>T</u>r.,
       DOMAIN
                                         GLU-RICH.
                            1010
                     321
 TT.
       DOMATN
                                         GLU-RICH.
                            2772
                    1846
                                         Missing (in isoform 2 and isoform 3).
 VT
       DOMAIN
                              28
                      17
 F'\Gamma
       VARSPLIC
                                         /FTId=VSP_004102.
 \mathbf{FT}
                                         QLQEEI -> LATRRD (in isoform 4).
                            1642
                    1637
       VARSPLIC
 \mathbf{F}\mathbf{T}
                                         /FTId=VSP_004103.
 ГŢ
                                         Missing (in isoform 4).
                             3911
       VARSPLIC
                    1643
 řΤ
                                         /FTId=VSP_004104.
 \mathbb{F}^{r_{L}}
                                         Missing (in isoform 3).
                             2182
                    2175
       VARSPLIC
  FT
                                         /FTId=VSP_004105.
  FT.
                                         SADTFQKVE -> Q (in isoform 5).
                             2183
                    2175
       VARSPLIC
  \mathbb{F}^{T}
                                          /FTId=VSP_004106.
                                         VFGFYNMCFSTLC -> GSSIPELAHSDAYQTREICSS
  FT
                             2907
                     2895
        VARSPLIC
                                          (in isoform 2, isoform 3 and isoform 6).
  FT
  FT
                                          /FTId=VSP_004107.
  FT
                                          Missing (in isoform 5).
                             2948
                     2895
        VARSPLIC
  FT
                                          /FTId=VSP_004108.
                                          STTQFHAGMRR -> ALSLTTSWQHHSARPTAPLFFEILSH
  FT
                     3901
                             3911
        VARSPLIC
  FT
                                          SLG (in isoform 6).
  \overline{c}^{\prime\prime}\Gamma
                                          /FTId=VSP_004109.
  \mathbf{F}^{\mathbf{T}}
                                          K \rightarrow KQ.
                             1347
        VERIANT
                     1347
  FT
                                          /FTId=VAR_010926.
  FT
                                          E \rightarrow Q (IN REF. 3).
                               76
                       76
        CONFLICT
  FT
                                          M \rightarrow I (IN REF. 3).
                              475
                      475
        CONFLICT
  FT
                                          E \rightarrow G (IN REF. 3).
                              554
        CONFLICT
                      554
  \mathbf{F}^{\mathbf{T}}
                                          R \rightarrow S (IN REF. 3).
                               638
                      638
        COMPLICT
  FT
```

```
N -> S (IN REF. 3).
               663
                      663
    CONFLICT
FT
                               H \rightarrow N (IN REF. 3).
               913
                      913
    CONFLICT
FT
                               K \rightarrow N (IN REF. 3).
                      956
               956
    CONFLICT
                               QKH -> PKP (IN REF. 1 AND 2).
FT
                      982
               980
    CONFLICT
                               Q \rightarrow P (IN REF. 1 AND 2).
FT
                      997
               997
    CONFLICT
                               Q \rightarrow P (IN REF. 1 AND 2).
FT
                     1001
              1001
    CONFLICT
FT
                               N \rightarrow D (IN REF. 3).
                     1020
    CONFLICT
               1020
                               V \rightarrow E (IN REF. 3).
                     1028
               1028
    CONFLICT
                               R \rightarrow P \text{ (IN REF. 1 AND 2)}.
FT
               1626
                     1626
    CONFLICT
FT
                               N \rightarrow T (IN REF. 3).
               1703
                     1703
     CONFLICT
FT
                               V \rightarrow G (IN REF. 3).
               1707 1707
     CONFLICT
\mathbf{F}\mathbf{T}
                               MISSING (IN REF. 5).
                     1803
               1802
     CONFLICT
FT
                               A \rightarrow P (IN REF. 3).
                     1843
               1843
     CONFLICT
FT
                         7.5%; Score 128.5; DB 1; Length 3911;
  Query Match
 Eest Local Similarity 20.1%; Pred. No. 3.3;
  Matches 77; Conservative 75; Mismatches 116; Indels 115; Gaps
          18 VKILKDNLAILEKQDKKTDKASEEVSKSLQAMKEILCGTNEKEPPTEAVAQLAQELYSSG 77
             QУ
          664 IEKLKDNLGIHYKQ--QIDGLQNEMSQKIETMQ-----FEKDNLITKQNQLILE---- 710
          78 LLVTLIADLQ--LIDFEGKKDVTQIFNNILRRQI------GTRSPTVEYISAHPHI 125
              QУ
          711 --ISKLKDLQQSLVNSKSEEMTLQI--NELQKEIEILRQEEKEKGTLEQEVQELQLKTEL 766
          126 LFMLLKGYEAPQIALRCGIMLRECIRHEPLAKIILFSNQFRDFFKYVELSTFDIASDAFA 185
                                             ΩУ
           -----NDLQEKFAQLEAEN-SILKDEKK 797
          767 LEKQMKEKE-----
          186 TFKDLLTRH -----KVLVADFLE-QNYDTIFEDYEKLLQSENYVTKRQSLKLLGELTL 237
              Talel I are the first of the left of the
 Q_{\mathcal{T}}
          798 TLEDMLKIHTPVSQEERLIFLDSIKSKSKDSVWEKEIEILIEENEDLKQQCIQINEEIEK 857
 Db
                                                     -YISKPENLKLMMNLLRD 264
          238 DRHNFAIMTK-----
                                                      1 | | : | | : | |
 QУ
          258 QRNTFSFAEKNFEVNYQELQEEYACLLKVKDDLEDSKNKQELEYKSKLKALNEELHLQRI 917
 Db
           265 KSPNIQFEA--FHVFKVFVASPHKTQPIVEILLKNQPKLIEFLSSFQKERTD-DEQFAD- 320°
                  ÔА
           918 NPTTVKMKSSVFDEDKTFVA---ETLEMGEVVEKDTTELMEKLEVTKREKLELSQRLSDL 974
 Db
           321 ----EKNYLIKQIRDLKK 334
  ŐА
                      975 SEQLKQKHGEISFLNEEVKSLKQ 997
  Db
```

Search completed: January 7, 2004, 16:45:28 Job time : 20 secs